

CCD OBSERVING AND DYNAMICAL TIME SERIES
ANALYSIS OF ACTIVE GALACTIC NUCLEI

By

ACHOTHAM DAMODARAN NAIR

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1995

May you have interesting data . . .

Ancient Chinese Curse

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Alex G. Smith, for his guidance, faith and vast patience. It would have almost impossible to perform the various interesting research projects without his encouragement. He has always had high expectations of me and I learnt from him the supreme importance of completing, at all costs, what I set out to accomplish. Finally, I thank him for all the useful conversations we have had over the years.

I would also like to thank Dr. Jose C. Principe for directing me through the intricate and fascinating world of neural networks. He always had time from his busy schedule to answer my questions, sometimes silly, arising from my always inquisitive mind. I also thank him for allowing me to use the computer facilities at the Computational NeuroEngineering Lab.

I thank my committee members, Dr. Malay Ghosh, for assistance with some of my early projects in statistics and very useful conversations; Dr. James H. Hunter, for discussions on the various physical models that may be applicable for AGN variability; and Dr. Robert J. Leacock, for assistance with graduate school matters and for serving on my committee.

I wish to thank my friends at the Dept. of Astronomy. Dr. Jaydeep Mukherjee for assistance in the Florida Space Grant proposal. To my deeply troubled friends Drs. Dave Kaufmann and J. C Liou for useful conversations and endless conversations/arguments over nothing. Dr. Sandra Clements for helping me learn the basics of observing, Dr.

Henry Kandrup for stimulating conversations and Dr. Ron Drimmel for assistance with the preparation of the dissertation.

Finally, the list would incomplete without mentioning office managers Debra Hunter, Carlon Ann Elton and Glenda Smith for their assistance in this tangled world of bureaucracy. Also to Drs. Henry D. I. Abarbanel, Eric Kostelich and Thomas Martin-Kruel for supplying software.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT	xii
CHAPTERS	
1. THE OBSERVING PROJECT	1
Introduction	1
History of CCDs	1
CCD : Theory and Working	2
Operation/s of the CCD at the 30" Telescope	4
Flat Fields, Dark Currents and Biases	8
Reducing the data	11
Photometry with IRAF	12
Need for extensive reduction?	15
Transformation to a Standard System	17
2. THE OBSERVING CAMPAIGN	20
Specific Results	20
Basic Techniques	22
The Lomb-Scargle Periodogram	24
3. NONLINEAR TIME SERIES ANALYSIS	27
Introduction	27
The Data	30
Basic Definitions and Techniques of Nonlinear Analysis	31
Dimensions	31
The Correlation Dimension	32
The Information Dimension	33
Dimension Calculations	35
Entropy and Redundancy	36
Linear Redundancy	38
Computation of Entropy in State Space	38
Results of Application	40
The "Test for Noise" Algorithm	43
Generating Surrogate Time Series with Identical Power Spectra	43
The Prediction Algorithm.	44
The Statistic	46

Application to Artificial Time Series	46
A Linear Stochastic Process	46
Sine curve with noise	47
Application to 3C 345	49
Nonlinear Stochastic Process?	53
Structure Function analysis	57
State Space Continuity Method	59
Application to Time Series	60
Local Linear Approximation Method	61
Application to time series	62
DISCUSSION AND CONCLUSION	65
4. DYNAMIC MODELING OF TIME SERIES USING ARTIFICIAL NEURAL NETWORKS	70
Dynamic Neural Networks	73
Training	77
Results	80
White Dwarf time series	80
3C 345 time series	87
Discussion	89
5. AGN TIME SERIES AND STOCHASTIC RESONANCE; A GENERIC MODEL FOR VARIABILITY	91
A Physical Model for Variability	92
Stochastic Resonance	93
Simulations	95
Discussion	99
6. SUMMARY	101
Summary	101
Future Work	102
BIBLIOGRAPHY	104
B.	111
BIOGRAPHICAL SKETCH	114

LIST OF TABLES

Table 1-1: Focii at the two different stations.	6
Table 1-2: Fiducial and computed magnitudes for ON 231	16
Table 3-1: Numerical values of z statistic for each embedding dimension.	57
Table 4-1: Parameters used for the results shown in figs. 4-6 and 4-7	83

LIST OF FIGURES

Figure 1-1: A schematic diagram of the "set up" of the CCD system and camera at RHO.	4
Figure 1-2: The quantum efficiency of the CCD chip to light of a range of wavelengths.	7
Figure 1-3: a) dark counts as a function of temperature for an integration time of 240 secs. b) for an integration time of 180 secs. c) Increase in dark counts as a function of time. The increase is linear. d) dark counts as a function of temperature for an integration time of 360 secs. $-45^{\circ}\text{C} = 0$, $-46^{\circ}\text{C} = 1$ and so on.	10
Figure 1-4: The color-magnitude difference diagram for M 67.	18
Figure 2-1: An intranight microvariability run for 3C 345 done in May 1993. In all three figures the top two series are that of the comparison stars, while that at the bottom is that for 3C 345.	21
Figure 2-2: Power spectra for the quasar 0735+178 (top) and BL Lac OJ 287 (bottom)	25
Figure 2-3: Structure functions for quasar 0735+178 (top) and BL Lac OJ 287 (bottom)	26
Figure 3-1: Top; the 3C 345 B-band time series after binning. Bottom; the original time series.	31
Figure 3-2: The correlation integral plot for 3C 345. The left-most curve corresponds to $d = 1$. The curves converge to a slope of 4.	36
Figure 3-3: Application of the redundancy test to the White Dwarf time series (Palus, 1994).	40
Figure 3-4: Redundancy and linear redundancy test to the AR2 (eqn 12) and sine curve (eqn 3.13)	41

Figure 3–5: The redundancy and linear redundancy test for the NSP (top). Notice the substantial difference between the behavior of the redundancy and the linear redundancy. Bottom, application to the OVV quasar 3C 345. Substantial similarities exist between the two sets.	42
Figure 3–6: Surrogate data created from the 3C 345 time series by a nonlinear transformation.	44
Figure 3–7: Top; a second order autoregressive process. Bottom; the z statistic as a function of time delay.	47
Figure 3–8: The z statistic as a function of time delay for the sine curve. The top-most curve in each fig. indicates a dimension of $d = 1$, the dimension d increases by 1 for successively lower curves. Top; no added noise. Notice the convergence for $d \geq 2$. Middle; for the same time series with 33% added noise. Bottom; same as above but with 99% added noise. .	48
Figure 3–9: Application of the “test of noise” algorithm to a simple non-stationary sine curve. Note that the z statistic does not display significant changes with respect to fig. 3–8, middle.	49
Figure 3–10: Application of the “test of noise” algorithm to 3C 345. For the upper curve $d = 1$. Notice that from about the third curve (that is, $d = 3$) onwards the curves seem to converge to a roughly fixed mean value. Top; for a prediction interval of 10. Middle; prediction interval of 20 and bottom; prediction interval of 30.	50
Figure 3–11: The general mutual information for dimensions 2 through 7 as a function of time lag for 3C 345. The top-most curve is for dimension 7 and decreases by 1 for successively lower curves.	51
Figure 3–12: Top; the z statistic for the first half of 3C 345. Bottom; the z statistic for the second half of 3C 345.	52
Figure 3–13: The time series generated by the nonlinear differential equation (3–18). .	53

Figure 3–14: The general mutual as a function of time lag for the nonlinear stochastic process. The top-most curve is for dimension 7 and decreases by 1 for successively lower curves.	53
Figure 3–15: Application of the “test of noise” algorithm to the nonlinear stochastic process. Top, for a prediction interval of 10. Middle, for a prediction interval of 20. Bottom, for a prediction interval of 30.	54
Figure 3–16: Top; the mean z statistic as a function of the embedding dimension for the nonlinear stochastic process. Bottom; the mean z statistic for 3C 345.	55
Figure 3–17: Top; the z statistic for the first half of the nonlinear stochastic process. Bottom; the z statistic for the second half of the nonlinear stochastic process.	56
Figure 3–18: Structure functions. a) SF of the time series generated by the NSP. b) for the first-differenced NSP. c) applied to 3C 345 and d) to the first-differenced 3C 345 time series.	58
Figure 3–19: The state space continuity method applied to the four time series.	60
Figure 3–20: Comparison of the prediction error for sine curve and the linear stochastic process using the Local Linear Approximation. In all the boxes the lower is for the sine curve.	63
Figure 3–21: The prediction error for the NSP and 3C 345. The curves for 3C 345 are the lower ones.	64
Figure 4–1: The NeuroSolutions bread board with the TDNN setup.	72
Figure 4–2: A schematic diagram of the Gamma NN	75
Figure 4–3: A scheme implementing teacher forcing.	78
Figure 4–4: The mean square error as a function of the number of iterations for the White Dwarf time series using the tapped delay-line memory. Training was stopped at 2600 iterations.	79

Figure 4-5: Power spectrum of the input white dwarf data.	80
Figure 4-6: Top, the White Dwarf time series. Followed by results of autonomous prediction after 1, 2 and 3 epochs using the tapped delay-line memory.	84
Figure 4-7: The White Dwarf time series, followed by the results of autonomous prediction after 1, 2 and 3 epochs using gamma memory.	85
Figure 4-8: Correlation Integral of the White Dwarf time series (left), of the tapped delay-line memory (center) and of the AP time series using the gamma memory (right).	85
Figure 4-9: Power spectra of the autonomously generated time series. Solid line is the time series from the TDNN and the dashed line from the Gamma NN.	86
Figure 4-10: Top, the original 3C 345 data. Successively lower figures are simulations using the gamma memory for 3 continuous epochs. The ordinate are normalized flux units.	88
Figure 4-11: Power spectra for 3C 345 (solid line) and the simulated time series.	88
Figure 5-1: Light curves of the OVV quasars 2345-167 (top) and 1156+295 (bottom)	92
Figure 5-2: Noisy time series generated by integrating eqn. 5-3. The noise variance increases from a) — c). Stochastic resonance can be seen for intermediate noise variance; that is, fig 5-2b.	96
Figure 5-3: Noise variance versus power for the parameters listed above. Indicates an optimum value of 1.2 at which SR occurs.	97
Figure 5-4: Simulated light curves generated by varying the noise variance.	98
Figure 5-5: Top, light curve of the quasar 3C 345. Bottom, simulated light curve obtained as a result of stochastic resonance.	98

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

CCD OBSERVING AND DYNAMICAL TIME SERIES
ANALYSIS OF ACTIVE GALACTIC NUCLEI

By

ACHOTHAM DAMODARAN NAIR

DECEMBER, 1995

Chairman: Alex G. Smith.

Major Department: Astronomy

The properties, working and operations procedure of the Charge Coupled Device (CCD) at the 30" telescope at Rosemary Hill Observatory (RHO) are discussed together with the details of data reduction. Several nonlinear techniques of time series analysis, based on the behavior of the nearest neighbors, have been used to analyze the time series of the quasar 3C 345. A technique using Artificial Neural Networks based on prediction of the time series is used to study the dynamical properties of 3C 345. Finally, a heuristic model for variability of Active Galactic Nuclei is discussed.

CHAPTER 1 THE OBSERVING PROJECT

Introduction

The Rosemary Hill Observatory (RHO), located at Bronson, Florida, has recently – around December 1992 – acquired the Photometrics Charge Coupled Device (CCD). CCDs are now the detector of choice for almost all astronomical observations. One of the main purposes of this instrument is to photometrically monitor the 200-odd Active Galactic Nuclei (AGN) that have been monitored photographically since December of 1968.

History of CCDs

The CCD concept was originated in 1970 by Boyle and Smith (1970). Astronomers were the first group of professionals to recognize the potential of the CCD for scientific imaging. The Jet Propulsion Laboratory (JPL) started a program to develop large array CCDs in 1973. JPL promoted interest in CCDs by demonstrating the capability of the CCD at many observatories worldwide. The enormous potential in imaging was instantly recognized and the interest in CCDs among astronomers became intense. Since the CCD is about 100 times more sensitive than a photographic film, new scientific discoveries were made at a rapid pace and generated new science. The CCD had become the sensor of choice and had revolutionized astronomy in much the same way as photographic film had 100 years ago. More data were available in much shorter periods of time and with the advent of larger telescopes one could see out to larger distances.

CCD : Theory and Working

The CCD works on the principle of the photoelectric effect; that is when photons strike certain types of material (like silicon or germanium) they release electrons. The incoming photons interact with the atoms of the material and excite the valence electrons into a conduction band. The number of electrons liberated is proportional to the number of photons of energy large enough (wavelength short enough) to overcome the surface work function. After this threshold, the number of electrons released is proportional to the illuminance times the exposure duration. The wavelength range over which this effect occurs is from soft x-rays to the near infrared.

A CCD consists of a rectangular array of pixels laid out in rows and columns and can be thought of as many shift registers composed of many pixel elements. The image forming section of the CCD is covered with closely spaced vertical registers or columns. A picture is read out of a CCD by a succession of shifts through the vertical registers. The device is serially read out line by line, pixel by pixel, representing the scene of photons incident on the device. The charge on the pixels is converted to an output voltage by an amplifier. An analogy may well explain things better. The CCD can be thought of as an array of buckets laid out row by row on a field. After one fills these buckets with water (say by rain) these buckets are transported, one row at a time, to a person who measures the amount of water in each bucket. Then if one were to plot on a graph a 2-dimensional portrait of the amount of water in each bucket, one would have a distribution of how much water there was in each part of the field (or how much rain fell in each part of the field). A one to one correspondence can be made from this picture to the CCD. The

drops of water (raindrops) are the photons, the buckets are the pixels, the transportation mechanism is the CCD shift registers. The sensitivity of the amplifier is expressed in volts per electron. The Star 1 CCD at RHO has 576 (rows) X 384 (columns) pixels, with each pixel being 22μ square. The quantum efficiency (QE) determines the sensitivity of the chip to photons incident upon it; this is a very important parameter in observing very faint objects. QE is a measure of photon loss and is wavelength dependent.

I was actively involved in getting the CCD “on-line” and streamlining its operation. Barring some minor additions and improvements to the logistics and “setup”, the CCD has been functional (and acquisition of AGN images, on a regular basis, has been in progress) for over two and a half years. The next and most important step was to find magnitudes and the errors associated for the objects. Each raw image obtained is a combination of photons from the objects of interest and background photons. The background photons are a source of noise. Noise is defined as anything that contributes to the uncertainty or errors of estimation. To get optimal results it is essential to know the sources of noise. A careful separation of the background photons from the objects of interest has to be done in the data reduction process. This reduction has a set of robust standard procedures which have been well documented (eg. see Howell, 1992). These reduction techniques have been primarily developed by the National Optical Astronomy Observatory (NOAO) and are called Image Reduction and Analysis Facility (IRAF). The treatment and removal of noise from an image involves a series of steps and is the key to successful photometry. I have streamlined this reduction process and provide a detailed exposition of all reduction steps in subsequent sections so that one can obtain magnitudes

for objects without much effort and time delay. However, one needs to experiment with parameters as much as possible.

A careful definition of the priorities and mode of operation is necessary since the intrinsic properties of a CCD are different from those of a photographic plate even though the final goal may be the same. This work has (should) make it possible for subsequent users at RHO to utilize the CCD technology with a minimum of effort and without repeating the time-consuming learning process.

Operation/s of the CCD at the 30" Telescope

The schematic diagram of the "setup" can be seen in fig. 1-1.

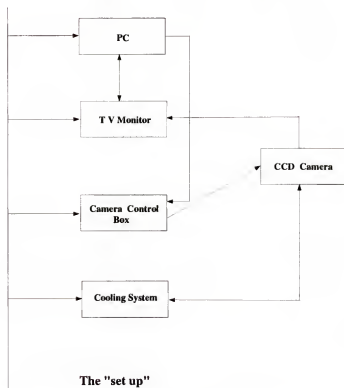


Figure 1-1: A schematic diagram of the "set up" of the CCD system and camera at RHO.

The main components of the system are a Personal Computer (PC), a 486 processor, a TV monitor and a tape back up system. The Photometrics Star_i.1.5 software has been installed on the PC and the PC is therefore in charge of all of the CCD operations. The coded commands from the software are received by the Camera Control Box and are interpreted and sent to the CCD camera head. The CCD camera head houses a 384 X 576 pixel CCD chip. The cooling system is an independent unit whose sole purpose is to supply the coolant to the camera head. The cooling system circulates a mixture of water and ethylene glycol and cools the CCD to a regulated temperature of -45°C . Both the Camera Control Box and cooling system are mounted on the telescope. The output image from the CCD camera is displayed on the TV monitor and a copy of the image is stored in the Camera Control Box. If desired this image may be stored on the hard disk of the PC in a binary image format, a TIFF format or a FITS format.

The following constitutes what must be done before the start of an observing run. First turn on the cooling system of the CCD, then turn on the CCD camera. Both of these operations can be accomplished by turning a switch on at the rear end of the telescope. The next step is to turn on the TV monitor and the PC that controls the operations. Go to the "Windows" menu and start up "Star_i.1.5". When the CCD temperature reaches -45°C it is ready to start exposing and obtaining images. The temperature may be left unregulated to get a cooler temperature (usually) depending on the ambient temperature. The CCD temperature apparently bottoms out at -63°C . Lowering the CCD temperature results in a substantial decrease in the dark current.

Before commencing operations with the CCD camera make sure that the air-conditioner inside the dome is turned off. The aim is to minimize the temperature gradient that may exist inside between the inside of the dome and outside it. The seeing (a term that describes the atmospheric turbulence along the line of sight) is sensitive to the temperature gradient and can produce poor quality images if substantial gradients are present. The field of view of the CCD camera is 10 X 15 arc-minutes. It may therefore be wise to take a short test exposure to see if there are an adequate number of comparison stars in the field of view.

There is a choice of using the "focus" mode of the software to arrive at a focus or it can be done manually. Suggestions for the focus for the four filters B, V, R and I are listed in table 1-1. The numerical values for the focus setting were arrived at after extensive experimentation and testing.

Table 1-1: Focii at the two different stations.

TABLE 1-1			
FILTER	STATION 2	STATION 3	Exposure time (secs)
B	87	165	1200
V	75	153	600
R	69	147	240
I	57	135	240
No filter	24	102	60

The values listed in table 1 are for “winter time” temperatures; that is for temperatures in the range of about 4°C (40°F) to 12°C (53°F). These values are temperature dependent and must be updated with temperature change. Go to lower values when warmer and to higher values when colder. It is usually difficult to obtain a good focus when the seeing is poor. The values of the foci listed in table 1 will change as a result of the recollimation that took place in September 1995. The relative offsets, however, should remain the same. The CCD chip is most sensitive in the red (700 nm). In terms of specific filter

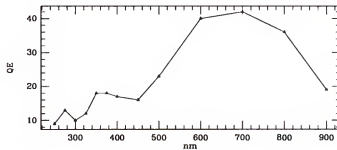


Figure 1-2: The quantum efficiency of the CCD chip to light of a range of wavelengths.

bands, it is most sensitive in the I, followed by R, V and B (see fig. 1-2). Signal-to-noise ratios (S/N) of an image are strongly dependent on the seeing and transparency. Some empirically determined exposure values under good seeing conditions and good transparency are listed in table 1 for a 15th magnitude object and a desired S/N of 40. The exposure times, naturally, increase by a factor of about 2.5 for successively fainter magnitudes. There can be substantial changes to the above estimates depending on the seeing and presence of clouds.

Flat Fields, Dark Currents and Biases

Obtaining a Flat Field

A CCD chip is a two-dimensional array of pixels. Any CCD chip will have sensitivity variations across its pixels due to its manufacturing process or its assembly. This means that each pixel does not exactly produce the same number of electrons per incident photon. Therefore this variation in sensitivity is a potential source of error. In order to normalize the sensitivity across the chip divide the image by a flat field. A flat field essentially is an image with almost uniform illumination and no stars; the purpose is to measure the response of each pixel to uniform illumination. There are a couple of ways of obtaining flat fields. The first is a dome flat field : illuminate uniformly a flat board located in the dome with light of constant intensity. Then, using exactly the same set up as that which was used to obtain images, make a flat field image. The second method is using the twilight sky either at sunset or at daybreak. One of the main disadvantages of using the second method is that illumination of the sky varies fairly rapidly with time. The average count of the image should be above 1000. Make flat fields for all filters used.

Obtaining a Dark Current

A dark current image is acquired to account for the offset (pedestal in counts). This is the most common type of additive noise. The most common type of dark current that is present in every image is thermally created electrons. Another source is locally "hot" pixels caused by unusually high rates of thermal electron creation. The dark current is a strong function of temperature and ideally we must get dark current images at all possible

operating temperatures. The common accepted way of measuring dark current is to take several exposures with the CCD camera shutter closed. Several such measurements are needed for each exposure time, and these can be averaged over each exposure time. The length of exposure of a dark current should be equal to that of the image. It may then be necessary to form a dark current database, where a set of dark currents for each exposure time and each temperature is averaged over at least 20 dark current images. A fiducial dark current for a particular exposure time and temperature may be formed by taking the median of the dark current images.

In fig. 3c a typical plot of the increase in the counts in analog-to-digital units (ADU) of the dark current as a function of integration time is shown. The increase is linear and is not a function of the temperature. The intercept on the ordinate is a function of the temperature.

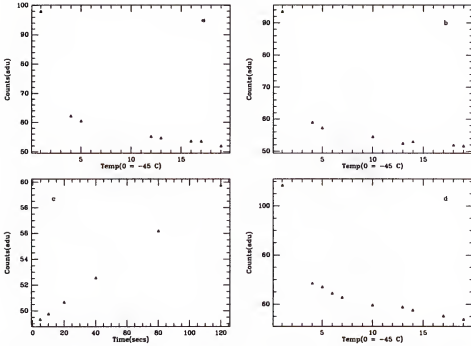


Figure 1-3: a) dark counts as a function of temperature for an integration time of 240 secs. b) for an integration time of 180 secs. c) Increase in dark counts as a function of time. The increase is linear. d) dark counts as a function of temperature for an integration time of 360 secs. $-45^{\circ}\text{C} = 0$, $-46^{\circ}\text{C} = 1$ and so on.

Fig. 3a, 3b, and 3d show the dark counts as a function of temperature for integration times of three, four and six minutes. Not surprisingly, the relationship between temperature and dark counts is nonlinear and can be fitted by an exponential law of the form

$$C = a \exp(-0.11T) + 50.1 \quad (1.1)$$

where C is the counts in ADUs and T is a scaled temperature with $-45^{\circ}\text{C} = 1$, $-46^{\circ}\text{C} = 2$ and so on; a is a scaling constant.

Bias

The bias is a zero-exposure offset measured in counts. It can be measured from fig. 1-3c as the value of the intercept on the ordinate.

Readout Noise

The initial charge measurement per pixel is an analog process; this is then passed to an analog to digital converter (A/D) to measure the charge in digital units. The charge level for each CCD pixel is measured by an output amplifier; the noise associated with this process is the readout noise. This is measured in electrons and places a lower limit on the charge that can be detected by a CCD. The readnoise does not depend on the geometry of the CCD.

Reducing the data

The working definition of a reduction process is steps taken to minimize the influence of data acquisition imperfections on estimation of the desired astronomical quantity. The first and foremost consideration is to examine the data to see if these defects exist on the image and to assess their impact on the photometry. One of the primary purposes of data reduction is to eliminate undesirable features on the image. The sources of some of the main undesirable features are inherent defects on the CCD chip, cosmic ray hits, hot spots and various types of noise. It is therefore wise to spend time to study the processes that have an impact on the data.

This means only perform reduction steps that matter. If the CCD chip (like the one at RHO) has no real gain or dark current variations and if the gradient across the field is negligible then any reduction that tries to correct for these effects will only degrade the final results (Gilliland, 1992). This implies that a considerable amount of time and effort must be spent on studying the effects of the various data reduction parameters and processes. Gilliland's (1992) rule of thumb is that if the application of reduction process results in a lower final noise, then that reduction process is appropriate, and vice versa.

Photometry with IRAF

Photometry in IRAF (Image Reduction and Analysis Facility) can be done with either of two packages, viz. APPHOT or DAOPHOT. It is recommended that one use DAOPHOT for crowded-field photometry. For AGN photometry APPHOT has been used throughout. Begin by setting the parameters in the four auxillary files that APPHOT uses; these are CENTERPARS, DATAPARS, FITSKYPARS and PHOTPARS. Each of the above 4 auxillary files has a list of parameters that have to be defined prior to commencing the photometry. There exist two types of parameters : a) the parameters that change from image to image (e.g. exposure time, zero point, etc.) and b) the parameters that are held fixed (e.g. minimum and maximum value of the data, aperture radius, etc.).

The parameters in the auxillary files are set to default values. Suggestions are provided in the IRAF manual on how to adjust the parameters depending on the kind of data. Discussed here are all of the parameters that have direct impact on photometry. In order to understand what parameters have a direct impact on the photometry it may be instructive to examine how IRAF computes the magnitude and error of a certain object.

The signal, which is the total counts in the object, is quantitatively defined as

$$S = \sum_{i,j}^{npix} C_{ij} - npix \cdot isky \quad (1.2)$$

where C_{ij} is the counts per pixel, $npix$ is the total number of pixels in the aperture and $isky$ is the background sky value per pixel. The summation is carried out over those pixels whose distance from the center of the object is less than a predefined radius. In magnitudes this signal may be expressed as

$$m = zpt - 2.5 \log_{10} S \quad (1.3)$$

where zpt is an arbitrary number used that may be adjusted to produce reasonable values for the magnitudes. The error in the magnitude is approximately equal to the inverse of the signal-to-noise ratio. The noise is computed as follows

$$N = \sqrt{err_1 + err_2 + err_3} \quad (1.4)$$

where N is the noise and

$$err_1 = npix \cdot \sigma^2 \quad (1.5)$$

where $\sigma^2 = \sqrt{\frac{sky}{gain} + \left(\frac{r_e}{gain}\right)^2}$, where sky is in counts, $gain$ is in electrons per ADU, r_e is in electrons and σ^2 is in counts. This term, err_1 , represents the noise in the sky background. The second term, err_2 , is defined as

$$err_2 = \frac{S}{gain} \quad (1.6)$$

This quantity measures the Poisson noise in the star image. Finally the third term, err_3 , in the expression for N above is

$$err_3 = \frac{npix^2 \cdot \sigma^2}{nsky} \quad (1.7)$$

where n_{sky} is the number of pixels used to measure the sky value. Extensive numerical schemes to reduce the noise (tasks in IRAF) may be employed; however an easy way to avoid these schemes is to use a larger integration time.

From the above, the error in the magnitude computed by IRAF is $SNR = \frac{S}{N}$; here SNR denotes signal to noise ratio. Note that this term is dependent on the gain.

For faint objects err_1 and err_3 dominate. For any given object the only way these can be reduced is to use the smallest aperture that gives reasonable results and/or increase the size of the sky annulus. The sky annulus must not be increased too much so as to get interference from nearby objects. A large aperture may add a lot of noise without increasing the signal to produce larger errors. But it has to be large enough that measurements made with it are independent of changes in seeing, telescope focus etc., that can occur during a night's observing. The choice of the "optimum" aperture with which to measure an object depends on its brightness. This value was found by trial and error to be 4.5 pixels and produced the smallest error in the photometry.

Gain and Readnoise

One of the first images that is taken at the telescope is the flat-field exposure and bias frame. From these we may compute the gain and readnoise, r_e , as

$$gain = \frac{(f_1 + f_2) - (b_1 + b_2)}{\sigma_{f_1 - f_2}^2 - \sigma_{b_1 - b_2}^2} \quad (1.8)$$

where f_1 and f_2 are two flat field images and b_1 and b_2 are bias frames. The two σ^2 s are variances of the resultant images after subtracting one flat field image from the other and one bias frame from the other. The choice of gain value depends on the science. The

gain is measured in electrons per count (ADU); in the Photometrics software the units are reversed, that is, ADUs/electron. A higher gain setting gives more dynamic range.

The readnoise is

$$r_e = \frac{gain \cdot \sigma_{b_1 - b_2}}{\sqrt{2}} \quad (1.9)$$

The readnoise is measured in electrons and places a lower limit on the the charge that can be detected by the CCD. It is the initial charge measurement per pixel as measured by the output amplifier, and the noise associated with that measurement is the readnoise.

Need for extensive reduction?

The issue of whether or not to perform basic image processing, subtraction of the dark current and division by a flat field, is considered here. This is best illustrated by an example. A 300-second V image of the BL Lac ON 231 was obtained on May 5th, 1994. Photometry on ON 231 was performed two ways using IRAF; first photometry without any reduction and second photometry with reduction.

There are four comparison stars available in the field of the image. One of these stars (#7) was used to set the zero point (eqn. 3). The results are shown in table 1-2.

Table 1-2: Fiducial and computed magnitudes for ON 231

Table 1-2				
ON 231, V, unprocessed				
Star	# 7	# 1	# 5	# 6
Computed comparison star magnitudes				
	15.60	12.10	14.82	13.19
Fiducial magnitudes				
	15.60	12.08	14.82	13.08
Error	-	0.02	0.0	0.11
Magnitude & error : 15.35 +/- 0.04				
ON 231, V, processed				
Star	# 7	# 1	# 5	# 6
Computed comparison star magnitudes				
	15.60	12.05	14.74	13.15
Fiducial magnitudes				
	15.60	12.08	14.82	13.08
Error	-	0.03	0.08	0.07
Magnitude & error : 15.26 +/- 0.06				

The V magnitudes for the comparison stars in the field of the OVV quasar ON 231, listed in table 1-2, were obtained from Wing (1973) and are referred to as fiducial magnitudes. The mean error was calculated by comparing the fiducial magnitudes with the computed magnitudes for the comparison stars. The star which was used to set the zero point of the photometry (in the above it was star #7) was excluded. The mean error is therefore $error = \frac{1}{N} \sum_{i=1}^N |m_f - m_c|$, where m_f is the magnitude of the fiducial star and m_c is that computed from photometry.

As can be seen from the above table, processing introduces a larger mean error. The origin of the larger mean error may be explained as follows. The standard deviation of the unprocessed image is 3.63 counts. After subtraction of the dark current and dividing by the flat field the standard deviation of the processed image rises to 4.21 counts. From the formulae (eqns. 4, 5 and 7) for computing above it can be seen that a larger standard deviation introduces larger noise, hence larger error. Most images taken with the RHO Star 1 CCD camera are relatively flat and smooth with some minor gradients present along the edges. This means that unless there are significant gradients present on the image one should not subtract the dark current as it tends to increase the noise in the data (Davis, 1995) and should be treated as as part of the local sky background. However, one should account for the pixel-to-pixel gain variations by dividing by the flat-field. Finally, a good rule of thumb is to operate in a regime of a high SNR (about 50) to alleviate problems arising from data reductions.

Transformation to a Standard System

The method adopted was that outlined by Henden and Kaitchuck (1990) and can be summarized as follows. There are two methods by which one can arrive at a transformation. The first method is to observe a collection of standard stars at the same zenith distance on a photometric night. The second method, considered superior to the first by Henden and Kaitchuck (1990), is to pick standard stars within a cluster close to the zenith. After performing the standard data reduction procedures, compute the V' magnitudes for a certain number of stars. Find the error by comparing them to the true magnitudes. Plot this error as a function of $B-V$ and do a linear regression. The slope

and the intercept of the line give us the transformation coefficients. Two images, B and V, for the open cluster M 67 were acquired on April 1st, 1994, a photometric night, with a 40-second exposure for B and a 10-second exposure for V. Figure 4 shows a typical plot obtained for the open cluster M 67. The coefficients for the slope and intercept are

$$V' - V = -0.023(B - V) + 0.634 \quad (1.10)$$

where V' is the computed value.

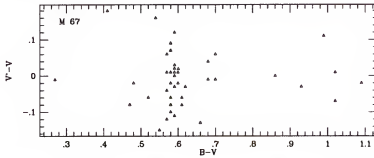


Figure 1-4: The color-magnitude difference diagram for M 67.

The linear correlation coefficient is only -0.01 . This is a typical result and has been the case for numerous other globular and open clusters such as M 3, M 92, M 71 and M72. The above result argues for having no transformation coefficients for the RHO Star 1 CCD chip. This has been borne out by the comparison of photometric magnitudes obtained without using transformation coefficients with magnitudes obtained at several observatories throughout the world. This can be seen in the following examples. A RHO R magnitude for 3C 345 for the night of August 4th, 1994 was 16.61 ± 0.03 and that obtained by Mark Kidger using the 0.82 m telescope at La Palmas for the same night was 16.60 ± 0.03 ; a second example is for BL Lac on the night of October 25th, 1994;

RHO R magnitude gives 14.34 ± 0.05 whereas that obtained by Tom Balonek using the 0.41 m telescope at Foggy Bottom Observatory was $R = 14.26 \pm 0.02$.

To conclude, on the basis of the above discussion the filters currently in use at RHO seem to closely match the pass bands of the Johnson BV system so that no significant transformation is required.

CHAPTER 2

THE OBSERVING CAMPAIGN

A collection of more than 200 Active Galactic Nuclei (AGN) have been monitored photographically from 1968 to 1992 at Rosemary Hill Observatory of the University of Florida. These AGN have been monitored photoelectrically with the Photometrics Star I CCD camera since 1993. These AGN are mainly radio-loud and they consist of BL Lacs, classical quasars and Seyfert galaxies. I participated in the ongoing program of monitoring the intensity variations of AGN. In addition to monitoring AGN I took part in several multiwavelength collaborative studies directed primarily at examining the evolution of large outbursts or flares across the electromagnetic spectrum. The aim of the multiwavelength studies are a) to obtain a multifrequency spectrum of a particular source, usually from the far UV to the radio, and to find the spectral slope in the UV-optical and optical-IR-radio regions. b) to monitor a flare event and follow its evolution in the optical and radio frequencies and to look for any lag in the rise or decay of the flux between the radio and the optical. I have also performed numerous microvariability studies on AGN. The term microvariability refers to variations of the order of seconds or minutes and in some cases hours.

Specific Results

Among the multiwavelength collaborative studies I took part were the 1991 outburst of the quasar 3C 345, the May-June 1992 outburst of the optically violent variable (OVV) quasar 3C 379. Even though the observations undersampled the flare events,

these endeavors have shown that the outbursts are probably due to a varying rate of electron injection into the jet. Also that the time lag between the rise of the optical flux and radio flux is zero. The flare spectrum can be represented by a power with a high negative index. There has also been some evidence that the inferred accretion rate varies with the jet luminosity (Webb et al., 1994).

As a specific case of a microvariability study consider the OVV quasar 3C 345. The microvariability time series is shown in fig 2–1. The study was conducted on the nights

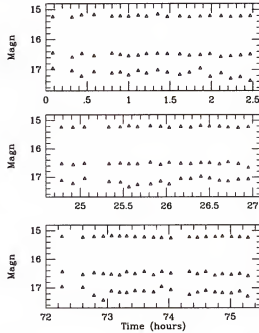


Figure 2–1: An intranight microvariability run for 3C 345 done in May 1993. In all three figures the top two series are that of the comparison stars, while that at the bottom is that for 3C 345.

of 21th (fig. 2–1, top), 22nd (fig. 2–1, middle) and 24th (fig. 2–1, bottom) May 1993, using an V filter with an exposure time of 480 secs. The quasar was relatively faint around 17.2 magnitude and 57 exposures were made over a span of about 3 hours each

night. The average signal-to-noise ratio was 25. Each image was flat-fielded to remove pixel-to-pixel variations. Two comparison stars, within 10 arc-minutes of the quasar, were chosen to contrast the variations of the quasar. In fig. 2-1 the top two series monitor the comparison stars and the bottom-most series shows the variations of the quasar. There appears to be two episodes of rapid change – microvariability. The first is at about 2 hours (fig. 2-1, top) where there is a brightening followed by decline. The whole event lasted about 30 minutes and the quasar underwent a change of 0.42 magnitude, from 16.95 to 17.37. The second episode was at 72 hours (fig. 2-1, bottom); that is, on the night of 24th, May 1993. Here again the quasar faded by about 0.47 magnitude, from 16.96 to 17.43, with the whole event lasting about 1 hour. Notice that during both the episodes the comparison stars have remained unchanged. No other clear cut evidence of microvariability can be seen in fig. 2-1. One way to confirm this microvariability would be to conduct another intranight study and look for similar variations. It is not clear what the source of these variations are; that is, whether these arise from hotspots on accretion disks or some instability on the jet. This issue is, however, addressed in chapter 5 in the context of stochastic resonance.

Basic Techniques

Here some basic time series analysis is conducted and certain statistical properties of a subset of the total AGN (about 210) are investigated. A large majority of the AGN display variations on time scales of minutes, hours, days, weeks and years. Due to the limitations imposed by the frequency of sampling (limitations due to bad weather, equipment malfunction and the object being close to the sun) one is largely restricted to

investigate variations of the order of years; this will henceforth be referred to as long term variations. Long term variations are essentially low frequency variations characterized by brightening and fading of the AGN over a period of several years. Most quasars and BL Lacs have, superimposed on these long term variations, sudden episodes of brightening and fading popularly referred to as flares; the flares typically last for weeks or months. One of the questions left unanswered, mainly due to data limitations, is whether the long term variations are the low frequency extensions of the flares. It has been suggested by Smith et al. (1993) that flares may be events quite different from the long term variations. This study is restricted to long term variations and in the next two sections basic time series analysis will be discussed. References to the papers and abstracts can be found in the appendix.

Two popular methods are discussed here, the unevenly sampled periodogram and the Structure function. In astronomy there are several factors that contribute to not being able to obtain data at fixed intervals. The sampling interval, especially at optical wavelengths, is dictated by weather conditions, state of the equipment and proximity of the object to the sun. As a result of these factors there inevitably are gaps in the data. Therefore, one may wish to fill the gaps in the data by interpolation and then use Fourier transforms or use the data without resorting to interpolation and use the Lomb-Scargle periodogram. Interpolation techniques are known to produce spurious results (Press et al. 1992); here no interpolation is performed and the Lomb-Scargle periodogram is used for spectral analysis.

The Lomb-Scargle Periodogram

If there are N data points sampled at discrete time t_i , $i = 1, 2, \dots, N$, the periodogram as a function of frequency ω is defined (Horne and Baliunas, 1986) as

$$P_N(\omega) = \frac{1}{2\sigma^2} \left\{ \frac{[\sum_{j=1}^N x(t_j) \cos \omega(t_j - \tau)]^2}{\sum_{j=1}^N \cos^2 \omega(t_j - \tau)} + \frac{[\sum_{j=1}^N x(t_j) \sin \omega(t_j - \tau)]^2}{\sum_{j=1}^N \sin^2 \omega(t_j - \tau)} \right\} \quad (2.1)$$

where τ is defined by

$$\tan(2\omega\tau) = \frac{\sum_{j=1}^N \sin 2\omega t_j}{\sum_{j=1}^N \cos 2\omega t_j} \quad (2.2)$$

and σ^2 is the variance of the data.

Horne and Baliunas (1986) arrived at a method of calculating a false alarm probability for a peak in the power spectrum. The false alarm probability is the probability that a peak of height z will occur assuming the data are pure noise. Suppose N_i independent frequencies were sampled and z is the height of a peak then the false alarm probability for this peak is given by $F = 1 - [1 - \exp(-z)]^{N_i}$; therefore, the probability that the given peak is real is $1 - F$. Horne and Baliunas (1986) empirically calculated the number of independent frequencies to be $N_i = -6.362 + 1.193N_o + 9.8 \cdot 10^{-4} N_o^2$

As examples consider the power spectra of the quasar 0735+178 (top fig. 1) and the BL Lac (bottom OJ 287). The power spectra were computed using eqn. 1 after subtracting a linear trend in the time series. In the case of 0735+178 the dominant peak is at 4.72 years with a false alarm probability of 0. For OJ 287 the dominant peak is at 11.3 years with a false alarm probability of 0.

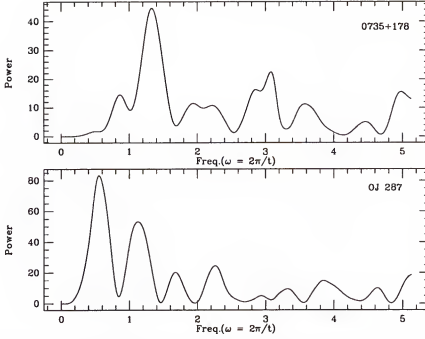


Figure 2-2: Power spectra for the quasar 0735+178 (top) and BL Lac OJ 287 (bottom)

The Structure Function

A second technique to probe the time series for time scales of variations is the structure function (SF) discussed by Simonetti et al. (1985) and applied to quasar time series by Smith et al. (1993). All data were analyzed using the first order SF. This can be defined as

$$SF(dt) = \frac{1}{N(dt)} \sum_{i=1}^N w(i) \cdot w(i + dt) \cdot [x(i + dt) - x(i)]^2 \quad (2.3)$$

where $N(dt) = \sum w(i) \cdot w(i + dt)$ and the weighting factor $w(i) = 1$ if a measurement exists in the i^{th} interval, 0 otherwise.

It can be seen from this equation that the SF does accept unevenly spaced data by assigning zero weights for intervals where there is no data.

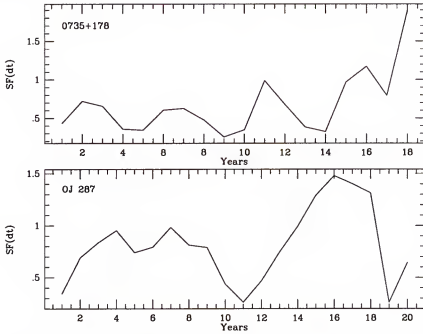


Figure 2-3: Structure functions for quasar 0735+178 (top) and BL Lac OJ 287 (bottom)

Fig. 2-2 shows the SFs for the above AGN, 0735+178 and OJ 287. The period of variation is the first minimum of the SF. In the case of 0735+178 the first minimum is around 5 years and for OJ 287 the first minimum is at 11 years. Both estimates are consistent with the power spectrum estimates of 4.72 years for 0735+178 and 11.3 for OJ 287. Details of computation of power spectra and SFs for 60 quasars can be found in Smith et al. (1993). A similar computation for BL Lacs, compact galaxies and radio-quiet quasars have been performed in Smith and Nair (1995).

CHAPTER 3

NONLINEAR TIME SERIES ANALYSIS

Introduction

We receive information from Active Galactic Nuclei (AGN) in the form of flux variations. One of the questions we would like to ask is : what do these data, accumulated over a certain period of time, tell us? To answer some parts of this question we look towards the theory of dynamical systems as a diverse and active field. Since the mathematical ideas and techniques have been well developed in this area, it is imperative for us to apply these concepts to AGN time series and to explore fully the information that may be hidden in the time series. There are many kinds of systems in nature : a) a system that produces pure random noise; b) a linear stochastic system; c) a nonlinear stochastic system; d) a chaotic system; and e) a deterministic system (neither chaotic nor stochastic). The first three listed here may be classified as non-deterministic systems, whereas the last two are deterministic systems. The working definition of determinism adopted in this paper is the popular notion that given a time series $x(n)$, there exists a map such that a future iterate $x(n+\tau) = f(x(n))$. Examples of the system in d) would be double pendula and for the system in e) would be a single pendulum (see Shinbrot et al. 1992). The systems in d) and e) may contain additive noise in the form of measurement noise.

A fundamental question is whether it is possible to extract useful information from experimental observations without placing any constraints on the kinds of phenomena that produce the time series we have obtained. In an analysis of the optical light curve of

the famous quasar 3C 273, Terrell and Olsen (1970) concluded that the light curve could be explained as originating from a system that produces an uncorrelated random process. In a recent analysis of the X-ray light curves of AGNs, Lehto et al. (1993), found that one source -- the Seyfert galaxy NGC 4051 -- may be produced by a low-dimensional (≤ 4.5) attractor. Vio et al. (1992) have suggested a "stochastic nature" for the X-ray light curve of the BL Lacertid 2155-304. Provenzale et al. (1994) used surrogate data methods to find that the 3C 345 time series may arise from a nonlinear stochastic process.

The first and necessary preliminary step in deducing the dynamics of a system from time series is to ascertain whether the time series originates from a nonlinear or a linear system. A system is said to be nonlinear if the differential equations used to describe the system are nonlinear. We try to show here how information-theoretic functionals, redundancies can be used for detecting nonlinearities in a time series. We define general redundancy and linear redundancy which is sensitive to linear dependence only. Thus a comparison of the general and linear redundancies can be used for a "qualitative" test for nonlinearity (see Palus et al. 1993). Detection of nonlinearity is based on the evaluation of redundancies of the time series and its delayed images in two ways : as a functional of the series probability distribution densities, to reflect nonlinearities and from the series covariance matrix to reflect linear relationships.

We are interested in determining the kind of system that generates the optical time series of the classical quasar 3C 345, which has been the subject of many extensive studies. Almost all of the studies that analyze the light curve have used Fourier Transform based techniques. In the present work we use short-term nonlinear predictability to

distinguish between the two broad categories; that is, deterministic or non-deterministic. This methodology is incorporated in the form of an algorithm, “test of noise”, that is capable of distinguishing a time series with a deterministic component from that generated by a) gaussian noise (white spectrum) b) linear stochastic noise (Kennel and Isabelle, 1992) and c) quite possibly from a nonlinear stochastic process. The test compares the nonlinear prediction error of the given time series with a set of gaussian time series that have the same average power spectral density as the given time series. A detailed description of the generation of the gaussian time series is given below. We have computed often-used fractal dimensions; that is, the Correlation dimension and the Information dimension, for the B-band light curve of 3C 345. The correlation dimension algorithm can indicate finite dimension even when the time series is from a stochastic process, as shown by Provenzale et al. (1994). Provenzale et al. (1994) indicated that stochasticity in AGN time series arises due to the unpredictable nature of the time series. However, neither the short-term predictability nor the long-term predictability of the time series has been tested by Provenzale et al. (1994) or by any other workers in the field. They have instead used surrogate data methods to test for determinism; these methods have drawbacks as shown by Theiler et al. (1992).

A technique employed to search for determinism in a time series is the State Space Continuity technique of Wayland et al. 1993. The test examines how nearest neighbors in state space move in time. If the movements are small it is evidence for determinism whereas large movements indicate stochasticity. Finally, a test first order approximation on nearest neighbors is the short-term prediction test using local linear approximations.

The quality of predictions, that is, the prediction errors, are used to differentiate between deterministic and stochastic systems. The main theme employed by the method is to find local linear approximations to the dynamics as originally suggested by Eckmann and Ruelle (1985) and applied to data by Farmer and Sidorowich (1987). This technique has been used as a method for filling gaps in data of a RV Tauri variable star by Serre et al. (1992).

Section 2 describes the data we have used. In section 3 we review the basic concepts and describe the techniques used. In section 4 we apply these techniques to artificial time series, section 5 discusses the application to the quasar 3C 345, and in section 6 the issue of whether the 3C 345 time series can arise from a nonlinear stochastic process is considered along with Structure Function analysis as in Provenzale et al. (1992). We discuss our results in section 7.

The Data

The data set consists of B magnitudes compiled mainly from the AGN monitoring program that has now been in place at Rosemary Hill Observatory (RHO), Univ. of Florida, for 26 years. Data prior to 1971 have been collected from various sources (Kinman et al., 1965; Barbieri and Abati, 1968; Lu, 1972). We have also made use of data after 1971 from the literature (Babadzanyants et al., 1975; Barbieri et al. 1977; Angione et al. 1981; Babadzanyants et al., 1985; Llyod 1984; Kidger and de Diego, 1990) to augment our database. The total amounts to about 1750 data points (see Fig. 3-1, bottom). The average error in the data set is 0.10 mag. Slightly less than 13% of the points were converted from PG magnitudes to B by using a correction of +0.26 mag.,

which has been empirically determined. We have followed the technique used by Vio et al. (1991) in reconstructing the time series for 3C 345. This comprises binning the data into, in our case, 7-day intervals and filling vacant bins by linear interpolation plus a random error that has the same mean as those from RHO (see Fig 3-1, top).

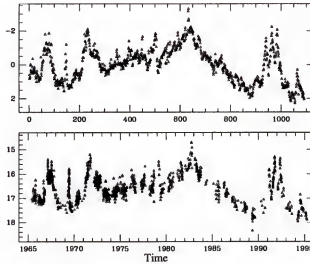


Figure 3-1: Top; the 3C 345 B-band time series after binning. Bottom; the original time series.

Subsequently, we subtracted out a small linear brightening trend of about 0.003 mag./year and have subtracted a base level magnitude.

Basic Definitions and Techniques of Nonlinear Analysis

Dimensions

Dimension is a fundamental quantity. Dimension can be conceived to be the information necessary to specify the position of a point on the attractor (an attractor is the set on which experimental points accumulate over infinite time) to within a given accuracy. It is the first level of knowledge necessary to characterize the attractor's

properties. It is also the lower bound on the number of variables needed in phase-space to model the dynamics. As examples, a stationary time-independent equilibrium – that is, a fixed point – has dimension zero; a stable periodic oscillation, a limit cycle, has dimension one; and a doubly periodic attractor (2-torus) has dimension two. The minimum dimension of the phase space into which the attractor may be placed is equal to $m_o = 1 + \text{int}(d)$ where d is the number of computed dimensions. The parameter m_o determines the number of active degrees of freedom. It can also be thought of as the number of dimensions required to resolve the attractor. There are many ways to define the dimension of a set. We discuss the two most often used.

The Correlation Dimension

Another type of dimension for experimental data or higher-dimensional dynamical systems, that is more practical to compute (than the capacity dimension d_c), is the correlation dimension d_g (Grassberger and Procaccia, 1983). The correlation dimension has been described lucidly in the literature (see for example Eckmann and Ruelle, 1985). The correlation dimension d_g measures how frequently certain regions of the attractor are visited.

Suppose that many points are scattered over a set. The typical number of neighbors of a given point will vary more rapidly with distance from that point if the set has high dimension. The correlation dimension may be computed from the correlation function $C(R)$ defined by

$$C(R) = \lim_{N \rightarrow \infty} [1/N^2 \sum_{i,j=1}^N H(R - |x_i - x_j|)] \quad (3.1)$$

where x_i and x_j are points on the attractor, $H(y)$ is the Heaviside function (1 if $y \geq 0$ and 0 if $y < 0$) and N is the number of points randomly chosen from the entire data set. The

Heaviside function simply counts the number of points within a radius R of the point denoted by x_i , and $C(R)$ gives the average fraction of points within R . The correlation dimension is defined by the variation of $C(R)$ with R :

$$C(R) \sim R^{d_{as}R} \rightarrow 0 \quad (3.2)$$

Therefore, the correlation dimension is the slope of a graph of $\log C(R)$ versus $\log R$. The line differs from a straight line for high and low values of R , for the following reasons. When R approaches the size of the phase space, $C(R)$ saturates at unity since all points are then included in the circle. Conversely, when R is smaller than the spacing between the data points, only one point lies in each circle and $C(R)$ levels off at $1/N^2$.

The quantities d_c and d_g are not equivalent. The capacity dimension depends only on whether small elements of phase space contain any points and does not take into account the differing numbers of points in the various elements. That is, small-scale variations of the density of points are ignored. On the other hand, the correlation dimension does include this effect.

The Information Dimension

A third quantity, the information dimension d_i , is related to the entropy. Like d_g , the information dimension depends on the distribution of points on the attractor. Suppose the attractor is covered by a set of n boxes of size ϵ and let the probability that a point is in the i th box be p_i . The metric entropy or missing information is

$$I(\epsilon) = -\sum_{i=1}^n p_i \log p_i \quad (3.3)$$

The information dimension d_i is defined by the equation

$$d_i = \lim_{\epsilon \rightarrow 0} [-I(\epsilon)/\log \epsilon] = \lim_{\epsilon \rightarrow 0} (\sum_{i=1}^n p_i \log p_i / \log \epsilon) \quad (3.4)$$

The variation of the entropy (or missing information) with ϵ scales as d_i . The basic argument for calculating d_i is as follows (Badii and Politi, 1985) : Draw k_1 points at random from the attractor without replacement. Pick another point x , called a reference point, and calculate non-negative distances from x to the k_1 points. Pick a number n between 1 and k_1 . Now let $\delta_n(x, k_1)$ denote the distance between x and its n th nearest neighbor. Continue by adding points at random and without replacement to the previous collection for a total of k_2 points. Record the distance again between x and its n th nearest neighbor based on k_2 points. This process continues for an increasing sequence of k_j 's until the distance between x and every other point has been calculated. Note that n is the order of the nearest neighbor. This is now repeated for a predetermined number of reference points $\{x\}_{i=1}^m$. If $\langle \delta_n(k) \rangle = \sum_{i=1}^m \delta_n(x_i, k) / m$ is the mean distance between the reference points and their n th nearest neighbors based on a sample of k other attractor points, then the nearest neighbor distances should scale as

$$\langle \delta_n(k) \rangle \sim k^{-1/d} \quad (3.5)$$

where d is the dimension of the attractor. In order to find the value of d , compute the distances $\langle \delta_n(k_j) \rangle$ for a number of values of n -- the order of the nearest neighbors -- and for k_j points.

The remarkable feature of these different definitions of dimension is that they have been shown to be related :

$$d_c \geq d_i \geq d_g \quad (3.6)$$

If the points are uniformly distributed on the attractor, the equality holds.

Are so many dimensions really needed to describe an attractor? Yes! Attractors contain more geometrical information than can be described by a single number. In much the same way, the mass distribution of an extended object requires an infinite number of parameters (the 'moments' of a mass distribution) for a complete characterization.

Since we have access to only one variable for our current problem -- that is, flux as a function of time -- we need to come up with a method of analyzing the real-world physical system (the AGN) which is multi-dimensional. It is, however, possible to analyze this multi-dimensional system on the basis of one scalar signal. This analysis can be carried out by embedding the given time series in an appropriately large embedding dimension (see, for example, Eckmann and Ruelle, 1985, or Lehto et al., 1993).

A third dimension, the capacity dimension, can be estimated from experimental data, using a box-counting technique. The capacity dimension is only concerned with the geometrical structure of the attractor. An excellent review of the capacity and other dimensions is given in Farmer et al. (1983).

Dimension Calculations

We have experimented with binning the data into many separate intervals to see the effect of binning on the computation of the dimension. Various bin sizes introduce a variation

in the computed dimension of about ± 0.35 . For bin sizes larger than 30 days, there are too few data points. We have calculated the correlation dimension d_g and information dimension d_i , for the time series of 3C 345. Fig. 3-2 shows a plot of $\log(C(r))$ versus

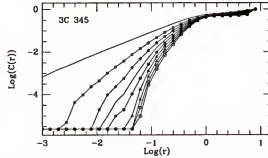


Figure 3-2: The correlation integral plot for 3C 345. The left-most curve corresponds to $d = 1$. The curves converge to a slope of 4.

$\log(r)$ for 3C 345. The curves seem to converge to a slope of 4.32; this slope is obtained from the approximately linear part of the curves for the large values of d at the right before they become asymptotic at large values of $\log(r)$. The reason for the deviation from a straight line for high and low values of R is mentioned above. The correlation dimension seems to agree with the information dimension, though it is not smaller than d_i , as alluded to above. We have computed a value of 4 for the information dimension d_i .

The correlation and information dimensions for noise were computed. A large dimension value usually is an indicator of noise. As expected, these computations give large values regardless of the time delays used.

Entropy and Redundancy

Every random variable x has an entropy or information $I(x)$. This is a non-negative number that roughly says how much information you would have if you knew the value of

x. Consider the two-dimensional case for the sake of simplicity. If we have two random variables x and y we can form a new random variable (x,y) . The distance between x and y is defined to be

$$d(x,y) = I(x) + I(y) - I(x,y) \quad (3.7)$$

This roughly tells you how much knowing x fails to tell you about y or vice versa. For example if $x = y$, the information in x is equal to the information in y , so that $d(x,x) = I(x) + I(x) - I(x,x) = 0$.

The mutual information states the general dependence of two variables. Consider a discrete source that produces a set of events $s_1, s_2, \dots, s_n$ with probabilities $P_s(s_1), P_s(s_2), \dots, P_s(s_n)$. A measure of how much "choice" is involved for event s to occur is the entropy H of the system,

$$H(S) = - \sum_i P_s(s_i) \log P_s(s_i) \quad (3.8)$$

that is the average information contained in the message.

We can compute the mutual information $I(S,Q)$ from two time series S and Q . Q is a time delayed image of S . $Q(t) = S(t + \tau)$ This is the definition for the mutual information :

$$I(S,Q) = H(S) + H(Q) - H(S,Q) \quad (3.9)$$

The mutual information $I(S,Q)$ is the sum of the information from the individual time series reduced by the information obtained from both of them. It measures the interdependence of two dynamical variables S and Q .

When there are n variables, $x_1, \dots, x_n$ (that is n delayed images of the time series) we get multidimensional mutual information called redundancy. Redundancy may be defined as a simple extension of the definition of mutual information for n variables $x_1, \dots, x_n$, as

$$R(x_1; \dots; x_n) = H(x_1) + \dots + H(x_n) - H(x_1, \dots, x_n) \quad (3.10)$$

Fraser (1989) has used marginal Redundancy to compute the metric entropy, which is the slope of the asymptotic marginal Redundancy curves for large n . A positive value for metric entropy indicates a chaotic time series. In this paper the use of redundancy is restricted to a qualitative test for the purpose of characterizing a time series as nonlinear or linear. The methodology adopted here is that outlined by Palus et al. (1993).

Linear Redundancy

From the $n \times n$ covariance matrix of $x_1, \dots, x_n$ variables linear redundancy may be defined as

$$R(x_1; \dots; x_n) = 1/2 \sum_{i=1}^n \log(c_{ii}) - 1/2 \sum_{i=1}^n \log(\lambda_i) \quad (3.11)$$

here c_{ii} are the variances and λ_i are the eigenvalues of the covariance matrix.

Computation of Entropy in State Space

Fraser and Swinney (1986) have outlined in great detail an algorithm to compute the mutual information, that is the 2-dimensional case. The first and most important step in computing the entropy is to compute the probabilities, according to eqn. 3-2, from discrete data. In order to accomplish this in the 2-dimensional case divide up the state space into four subregions. Then count the number of points present in each cell and

calculate the probability $P = \frac{N_{cell}}{N_{total}} \delta n_1 \dots \delta n_m$, where N_{cell} is the number of points in a particular cell, N_{total} is the total number of points and δn is the size of the individual cell. For computational efficiency we need to divide only the state space region that contains the attractor. The state space is divided recursively until a desired resolution or no further substructure is seen. The algorithm used in this paper is the algorithm described by Fraser and Swinney generalized to m dimensions. This extension to more than two dimensions is done by dividing up the state space into 2^m subregions. Greater details of the algorithm can be found in Fraser and Swinney (1986) and a wide range of applications have been listed in Fraser (1989).

Eigenvalues for calculating linear redundancies were done using the subroutine JACOBI from Press et al. (1992). Typically, we plot the redundancy and linear redundancy as a function of lag (in units of data samples). The different curves in each figure correspond to the number of dimensions increasing from bottom to top. The test consists of comparing the redundancy (in bits) with the linear redundancy. If the shapes of the curves as functions of lag are the same then the time series is considered to originate from a linear process. If, however, there are large dissimilarities in their shapes then the time series may arise from a nonlinear system, meaning that there is a nonlinear relationship between the variables (Palus et al., 1993). The qualitative differences can be defined as follows. Specifically, one looks for consistent positions of minima of time lag between the linear redundancy and redundancy plots, the dependence of redundancies as a function of lag and the rate at which the linear redundancy decreases to values close to zero compared with that of the redundancy.

Results of Application

The results of a test conducted on Data Set E for the 13th continuous segment of the Santa Fe Time Series Prediction and Analysis Competition (Clemens, 1994) are shown in fig 3-3.

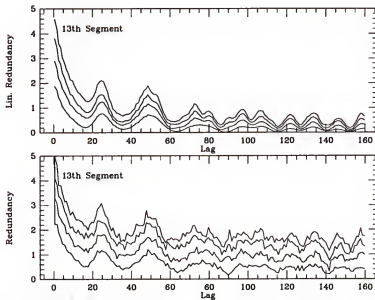


Figure 3-3: Application of the redundancy test to the White Dwarf time series (Palus, 1994).

Only the shapes of the curves and not their numerical value are of interest. A comparison of the shapes of the curves for dimensions 2 (lower most) through 5 (top most) for the linear redundancy plot as well as the redundancy plot show them to be approximately the same. This indicates that the time series is linear. These results corroborate with that obtained by Palus (1994) with minor variations in their shapes. The values of the redundancies are in bits.

Next consider the application to two artificially generated time series. The first is a second order autoregressive time (AR2) series generated by

$$x(t) = a_1x(t-1) + a_2x(t-2) + r \quad (3.12)$$

where r is white noise with mean zero and variance one. The parameters used were $a_1 = 1.3$ and $a_2 = -0.4$ (shown in fig. 3–7, top). The second is a quasi-periodic sine curve of the form

$$\sin(\omega_1 t) + 0.1 \sin(3\omega_1 t) + \sin(\omega_2 t) + 0.2 \sin(3\omega_2 t) + r \quad (3.13)$$

with $\omega_1 = 0.1$ and $\omega_2 = 0.3$, while r is white noise with mean zero and variance one. Fig. 3–4 top left and fig. 3–4 top right show the case for the linear stochastic process.

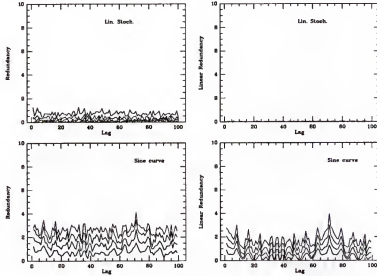


Figure 3–4: Redundancy and linear redundancy test to the AR2 (eqn 12) and sine curve (eqn 3.13)

Not surprisingly, both the redundancy values are approximately zero; that is, the variables are uncorrelated or that there is almost zero autocorrelation in time series; we

cannot learn anything about the future state of the system knowing its present state. The situation for the quasiperiodic sine curve with a large noise content is almost opposite to that for the linear stochastic process. Here the numerical values are significantly above zero and more importantly, the shapes of the curves are almost the same (fig. 3–4 bottom left and 3–4 bottom right). This indicates linearity of the time series.

In the case of the time series of the nonlinear stochastic process (NSP) there is a marked difference in terms of the plots (fig. 3–5 top left and 3–5 top right) for the redundancy and linear redundancy. The plots for the 3C 345 time series (fig. 3–5 bottom left and 3–5 bottom right) to a large extent resemble that of the NSP.

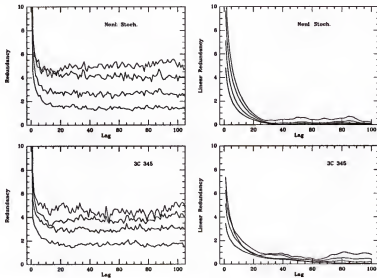


Figure 3–5: The redundancy and linear redundancy test for the NSP (top). Notice the substantial difference between the behavior of the redundancy and the linear redundancy. Bottom, application to the OVV quasar 3C 345. Substantial similarities exist between the two sets.

There are two clear distinctions that can be made. Firstly, the redundancy has substantial structure; that is, the dependence of redundancies on lag, whereas the linear

redundancy detects almost no structure, with the mean values for the redundancy being substantially higher than those for the linear redundancy. Secondly, the descent of the linear redundancy is much faster than that for the redundancy and the linear redundancy is close to zero for lags greater than 10. The two factors together argue for a nonlinear description for the NSP and the 3C 345 time series. One could further say that because of the close resemblance of the plots for the redundancy and linear redundancy in the cases of the NSP and 3C 345, the 3C 345 time series is consistent with that of a NSP.

The "Test for Noise" Algorithm

Kennel and Isabelle (1992) have devised a test to distinguish a non-stochastic time series from high-dimensional linearly correlated noise. We use that test here to determine whether the time series has a deterministic component or if it comes from linearly correlated noise or pure noise. Kennel and Isabelle use short-term predictability as a tool to distinguish between a deterministic process and randomness. They compare the prediction error of the given time series to the prediction errors, computed identically, of a set of random time series which have the same average power spectral density and thus the same autocorrelation as the original data series. If the prediction errors are sufficiently smaller than those of the random series we may reject the null hypothesis that the input times series is noise.

Generating Surrogate Time Series with Identical Power Spectra

The procedure adopted in the "test of noise" algorithm to create the null hypothesis follows "Algorithm II" of Theiler et al. (1992), also referred to as the "histogram

transformation” by Kennel and Isabelle (1992). Begin by generating a set of random numbers with a gaussian distribution equal to the length of the original time series $x(t)$. Rearrange the random numbers into a time series $R(t)$ whose rank orders follow the same sequence as $x(t)$. This means that, if $x(1)$ is the ninth highest amplitude of $x(t)$, then $R(1)$ should be the ninth highest random number from the set. Now find the random phase surrogate $P(t)$ (a random phase surrogate may be obtained by randomizing the phases of the Fourier Transform, see “Algorithm I” of Theiler et al. (1992)) of $R(t)$ and replace each value of $R(t)$ by its corresponding amplitude from $x(t)$ to get a new time series $G(t)$. This series may have different spectral properties than the original $x(t)$; however the underlying gaussian processes $R(t)$ and $P(t)$ have identical power spectra. Therefore $G(t)$ assumes that $S(t)$ was a manifestation of a linear stochastic process that was passed through a nonlinear filter. The surrogate time series created through this nonlinear transformation can be found in fig. 3–6 and is visibly different from the 3C 345 data.

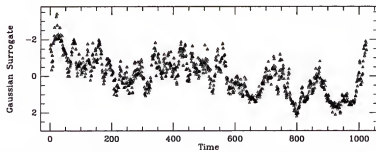


Figure 3–6: Surrogate data created from the 3C 345 time series by a nonlinear transformation.

The Prediction Algorithm.

This algorithm relies on the standard time-delay embedding for reconstructing the phase space. Given an input scalar time series $x(n)$, with integer “time” $n = 1 \dots N$, and with

an appropriate choice of T (usually obtained from the mutual information criterion) and d (obtained generally from the correlation integral), one forms vectors in d -dimensions (Takens, 1981) :

$$y(n) = [x(n), x(n+T), \dots, x(n+(d-1)T)] \quad (3.14)$$

The integer quantities T and d are the time-delay and embedding dimension, respectively. The prediction problem is : given an input vector $\mathbf{z}$ belonging to $\mathbb{R}^d$ that represents the current position in phase space, predict the next scalar component τ time steps ahead of the last component of $\mathbf{z}$. Local information about the future evolution of the close neighbors of $\mathbf{z}$ is used to predict the future evolution of $\mathbf{z}$. Specifically, we find the single closest neighbor $y(\eta)$ to the reference point $\mathbf{z}$. The predicted scalar, given a fixed prediction interval $\tau > 0$, is $x(\eta + (d-1)T + \tau)$, the future iterate of the nearest neighbor to $\mathbf{z}$.

A “k-d” tree search algorithm is used to accomplish rapidly the search for nearest neighbors (one may wish to use a brute force method to accomplish this in order to avoid complicated programming, at the expense of consuming longer CPU time). This formula is applied to every vector of the data set $y(j)$ to obtain a set of prediction errors. At every point $y(j)$ we compute the error $e(j)$ involved in predicting the next step forward, by first finding the nearest neighbor in the data set of $y(j)$, denoted by $y(\eta_j)$, and computing the difference between the neighbor’s future iterate and the actual evolution of $y(j)$:

$$e(j) = |x(\eta_j + (d-1)T + \tau) - x(j + (d-1)T + \tau)| \quad (3.15)$$

With N scalar data points, one can compute predictions for $N - (d-1)T - \tau$ vectors.

The Statistic

The whole aim of the test is to compare the sizes of the prediction errors (pe) of the true data set, $e(j)$, to the errors on the faked data sets, e^f , $r = 1 \dots M$. Now one can ask whether the pe from the true data set and from the faked data sets come from the same underlying distribution, or are the pe for the true data set statistically smaller than those of the pe for the faked data sets? The null hypothesis is that the true (or input) data set is pure noise. A quantitative way to find this out is to use the Mann-Whitney rank-sum statistic

$$U = \sum_{i=1}^{N_2} \sum_{j=1}^{N_3} H(A_i - B_j) \quad (3.16)$$

where N_2 is the number of elements in A, $N_3 = MN_2$, the number of elements in B, and $H(*)$ is the Heaviside step function : $H(x) = 1$ for $x > 0$ and $H(x) = 0$ for $x \leq 0$. For large N_2 , the quantity

$$z = \frac{U - N_2 N_3 / 2}{\sqrt{(1/12) N_2 N_3 (N_2 + N_3 + 1)}} \quad (3.17)$$

is normally distributed with zero mean and unit variance, under the null hypothesis that the two observed samples came from the same distribution. If on the other hand, the samples in set A were smaller than those of B, in a statistically significant way, we would observe small values of z . Thus, for example, if we observe a single z value less than -2.33 , we can state that we disprove the null hypothesis at the 99% confidence level.

Application to Artificial Time Series

A Linear Stochastic Process

We applied the "test of noise" algorithm to the second order autoregressive time series (eqn. 3-12); the time series is shown in fig. 3-7 top. Fig. 3-7 bottom shows

the z statistic as a function of time-delay and is plotted for 8 embedding dimensions. Notice that the z statistic for $d = 1$ to $d = 4$ remains positive to slightly negative. For

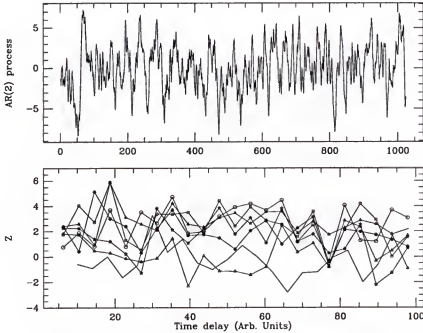


Figure 3-7: Top; a second order autoregressive process.
Bottom; the z statistic as a function of time delay.

higher dimensions, the statistic takes on more negative values; however, the overall mean z statistic for a particular value of d is around 0.

Sine curve with noise

Figure 3-8 shows the application to the quasi-periodic sine curve of eqn. 3-13. We first ran the algorithm with no noise added (fig. 3-8, top), then with 33% added noise (fig. 3-8, middle) and finally with 99% added noise (fig. 3-8, bottom). Application of the “test of noise” algorithm to a quasi-periodic time series. The top-most curve, in each figure, is for dimension 1 and the dimension increases by 1 for successively lower curves.

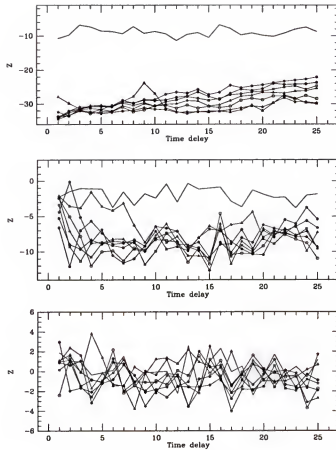


Figure 3-8: The z statistic as a function of time delay for the sine curve. The top-most curve in each fig. indicates a dimension of $d = 1$, the dimension d increases by 1 for successively lower curves. Top; no added noise. Notice the convergence for $d \geq 2$. Middle; for the same time series with 33% added noise. Bottom; same as above but with 99% added noise.

There are a couple of points of interest. Notice first that the curves converge for $d = 2$, and larger, which indicates the presence of 2 variables in the system; second, note that the z have large negative values even with 99% added noise, but the reason for this is that the signal is well-defined and deterministic. The relatively large negative values of the z statistic indicate that the time series generated using the quasi-periodic sine curve with added noise is definitely deterministic.

To test how non-stationarity of the time series affects the algorithm, we added a linear term to the above equation for the quasi-periodic sine curve to make it $\sin(\omega_1 t) + 0.1 \sin(3\omega_1 t) + \sin(\omega_2 t) + 0.2 \sin(3\omega_2 t) + a \cdot t + r$, with $a = 0.5$, ω_1, ω_2 taking on the values listed above and 33% added noise. Fig. 3-9 shows the performance of the algorithm; it

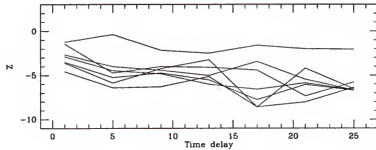


Figure 3-9: Application of the “test of noise” algorithm to a simple non-stationary sine curve. Note that the z statistic does not display significant changes with respect to fig. 3-8, middle.

can be seen that the z statistic is similar to fig. 3-8 (middle) with minor variations due to the addition of the non-stationarity term. This makes us believe that non-stationarity has almost no effect on the z statistic, but may induce minor variations on the shapes of the curves.

Application to 3C 345

Figs. 3-10 top, 3-10 middle and 3-10 bottom, show the application of the test of noise algorithm to the time series of 3C 345. The figures show the z statistic as a function of time delay and embedding dimension. We experimented with numerous prediction

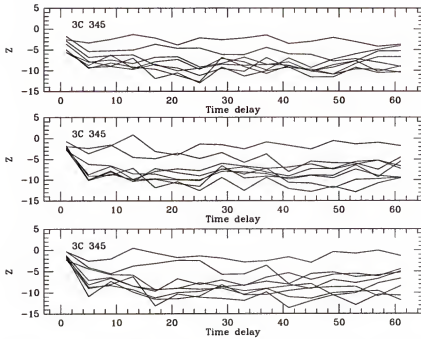


Figure 3-10: Application of the “test of noise” algorithm to 3C 345. For the upper curve $d = 1$. Notice that from about the third curve (that is, $d = 3$) onwards the curves seem to converge to a roughly fixed mean value. Top; for a prediction interval of 10. Middle; prediction interval of 20 and bottom; prediction interval of 30.

intervals; the figures display the statistic for three typical prediction intervals : 10, 20 and 30. The first minimum of the general mutual information was chosen as the time delay τ . The general mutual information, which measures the general dependence of two or more variables, is considered to be a more appropriate choice for picking a time delay τ than the autocorrelation function (Fraser and Swinney, 1986). The general mutual information can be loosely defined as a “nonlinear autocorrelation” function. The general mutual information as a function of lag is depicted in fig. 3-11 for dimensions 2 through 7 with the top-most curve being for dimension 7 and decreasing by 1 for successively lower curves. From fig. 3-11 we see that the first minimum of the general mutual

information, for dimensions 5, 6 and 7, is around 65 and this is the value we have picked for τ . This value for τ is smaller than value that obtained using the autocorrelation function which is about 80.

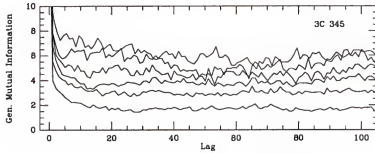


Figure 3-11: The general mutual information for dimensions 2 through 7 as a function of time lag for 3C 345. The top-most curve is for dimension 7 and decreases by 1 for successively lower curves.

Each of the various curves we see in the figures corresponds to an embedding dimension; six embedding dimensions were considered. There seems to be ample evidence to show that the time series is not generated by a linear stochastic process. The z values seem to drop to their lowest around $d = 4$; the mean z statistic is much below -2.33 (the 99% confidence level); this indicates that the hypothesis of the time series being a linear stochastic process can be firmly rejected. This may be one of the first hints that the underlying physical process is deterministic. However, as we discuss in the next section, the argument of Provenzale et al. (1994) that the 3C 345 time series may arise from a NSP cannot be firmly rejected. Although there is some measurement noise (i.e. observational error), we can conclude that the system that generates the time series has a fairly large noise content (large amount of intrinsic noise). We arrive at this conclusion based on fig. 3-8, where the addition of noise tends to drive the z

values to smaller negative values. However, no quantitative value can be assigned to the noise content in the 3C 345 time series, nor can the algorithm distinguish between measurement noise and dynamical noise.

To get a handle on how truncation of data affects the z statistic we split the time series into two halves and ran the “test of noise” algorithm on the two halves separately. Fig. 3–12 (top) shows the behavior of the z statistic on the first half and fig. 3–12 (bottom) shows its behavior on the second half. There is almost no change except, naturally, for a small increase in the mean value of the z statistic. Extensive tests were done by Smith et

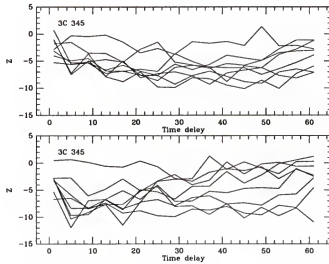


Figure 3–12: Top; the z statistic for the first half of 3C 345.
Bottom; the z statistic for the second half of 3C 345.

al. (1993) to show that the effects of data truncation do not significantly affect the power spectrum as long as the data run exceeds the time scale indicated by the power spectrum. This division into halves has been done to NSP time series in the following section.

Nonlinear Stochastic Process?

The primary contention of Provenzale et al. (1994) is that the optical light curve of 3C 345 may be mimicked by a NSP. The time series from this NSP is generated by integrating the following nonlinear differential equation :

$$\frac{dy(t)}{dt} = (\alpha - 0.5)\beta - y(t) + [2\beta y(t)]^{\frac{1}{2}}w(t) \quad (3.18)$$

where $w(t)$ is a gaussian white noise process. We have obtained a time series (see fig. 3–13) from Vio (1995) integrated with the same parameters and time step as that used in Provenzale et al. (1994). Linear interpolation of the artificially-created gaps was

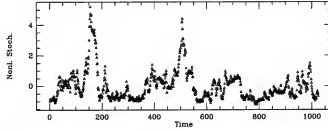


Figure 3–13: The time series generated by the nonlinear differential equation (3–18).

performed at the same epochs as those for the 3C 345 time series. Again, the general mutual information was chosen to select a value for the time delay τ . The general mutual information does not show a minimum as it did for 3C 345 (see fig. 3–14).

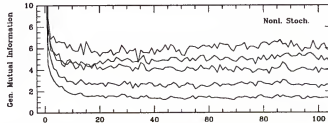


Figure 3–14: The general mutual as a function of time lag for the nonlinear stochastic process. The top-most curve is for dimension 7 and decreases by 1 for successively lower curves.

Therefore we selected the same value of 65 as we did for 3C 345. In figs. 3–15 top, 3–15 middle and 3–15 bottom, the results of application of the test of noise algorithm for three different prediction intervals and 8 embedding dimensions are displayed.

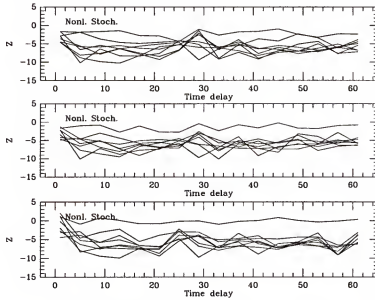


Figure 3–15: Application of the “test of noise” algorithm to the nonlinear stochastic process. Top, for a prediction interval of 10. Middle, for a prediction interval of 20. Bottom, for a prediction interval of 30.

Clearly the algorithm does reject the null hypothesis (the input time series is correlated noise), as can be seen from negative values for the z statistic, and does indeed recognize the NSP as a deterministic process. However, the behavior of the z statistic is obscured due to the clustering of the curves. A close look at the mean z statistic given in Table 1 shows that the z values for 3C 345 are consistently lower than those for the NSP; that is the 3C 345 time series shows more predictability than the NSP. In fig. 3–16 we examine the mean z statistic as a function of the embedding dimension for 3C 345 and the NSP. Fig. 3–16 (bottom) shows a tendency for the mean z statistic of the 3C 345 time series to

fall to its minimum value for a dimension of 4 with a modest rise thereafter. The behavior of the mean z statistic of the NSP time series is markedly different (fig. 3–16, top) in that the mean z statistic does not point to a clear minimum for dimensions beyond 3.

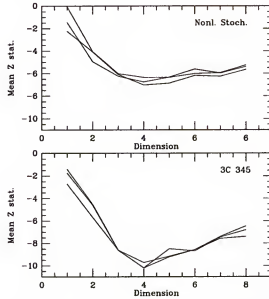


Figure 3–16: Top; the mean z statistic as a function of the embedding dimension for the nonlinear stochastic process. Bottom; the mean z statistic for 3C 345.

This, the mean z values and behavior of the z statistic, is an indication that the two time series do not originate from the same process. However, since we lack a quantitative criterion to establish the significance of lower z values and the behavior of the z statistic as a function of the embedding dimension, we are hesitant to unambiguously rule out the possibility that the 3C 345 time series may be from some NSP.

Data truncation has only minor effects on the NSP time series and drives the z statistic to slightly larger values. This is similar to the effect observed for 3C 345. Fig. 3–17 (top) displays the z statistic for the first half of the NSP and fig. 3–17 (bottom) for the

second half. The minor changes in shapes cannot be assigned any significant meaning in the case of the NSP as well as for 3C 345.

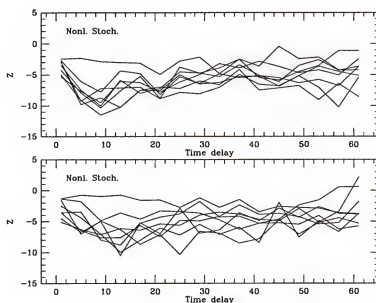


Figure 3–17: Top; the z statistic for the first half of the nonlinear stochastic process.
Bottom; the z statistic for the second half of the nonlinear stochastic process.

Table 3-1: Numerical values of z statistic for each embedding dimension.

TABLE 3-1			
Dim.	Mean z		
3C 345	Prediction Interval		
	10	20	30
1	-2.71	-1.74	-1.41
2	-5.68	-4.60	-4.52
3	-8.60	-8.63	-8.60
4	-9.72	-10.19	-10.20
5	-9.16	-8.49	-9.21
6	-8.57	-8.69	-8.57
7	-7.57	-7.45	-7.39
8	-7.39	-6.81	-6.46
Nonl.			
Stoch.			
1	-2.25	-1.46	-0.06
2	-4.08	-4.95	-4.06
3	-6.08	-6.24	-6.01
4	-7.02	-6.75	-6.35
5	-6.86	-6.33	-6.35
6	-6.19	-5.62	-6.02
7	-6.26	-5.97	-5.93
8	-5.65	-5.39	-5.25

Structure Function analysis

Provenzale et al. (1992) have shown that the Structure Function (SF) is one of the more appropriate and reliable methods for time series analysis. It has been shown to be

able to distinguish between a NSP and a deterministic one (in their case, the familiar Lorenz system). Further, for a fractal signal the SF has the property

$$S(n) \approx n^{2H} \quad (3.19)$$

applicable for small values of n , where H is a scaling exponent which is about 1 (see Provenzale et al., 1992 for more details). This method is unfortunately absent in their more recent paper (Provenzale et al. 1994). We have applied SF analysis to the time series of 3C 345 and the NSP. Fig. 3–18 illustrates the result of this application.

Fig. 3–18a is the SF of the data supplied by Vio. Fig 3–18b is the SF of the first difference of the same times series. Figs. 3–18c and 3–18d are the SFs of the 3C 345 time series and its first difference. Both time series display some scaling behavior and

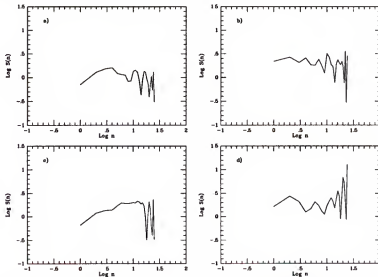


Figure 3–18: Structure functions. a) SF of the time series generated by the NSP. b) for the first-differenced NSP. c) applied to 3C 345 and d) to the first-differenced 3C 345 time series.

in both cases the scaling regime seems to have vanished for the differenced time series. There does not seem to be any other significant difference discernible in the SF for 3C

345 or the NSP data either in the differenced or non-differenced versions. The scaling exponent H (in the notation of Provenzale et al., 1992) for small values of n for the 3C 345 time series is 1; it is about the same for the NSP data. Even though there are some differences between the between the SF for the NSP and that for 3C 345, the differences are not as clear-cut as they are in Provenzale et al. (1992). Since there are no significant differences in the SFs for both data sets the results are ambiguous.

State Space Continuity Method

The state space continuity technique was first described by Wayland et al. (1993) and the methodology followed in this paper is almost identical to that of Wayland et al. (1993). After constructing a vector time series according to eqn. 7 pick an arbitrary vector x_0 on the attractor. Find a predetermined number, k , of NNs to this reference vector x_0 , x_1 , ..., x_k . Then find images of these k NNs, y_1 , ..., y_k , by projecting them through a certain fixed horizon, or translation, T . Compute the difference between the NNs (of x_0) and their images

$$v_i = y_i - x_i \quad (3.20)$$

and find a mean difference

$$\langle v \rangle = \frac{1}{k} \sum_{i=1}^k v_i \quad (3.21)$$

Using the terminology of Wayland et al. (1993), compute a normalized translation error as

$$e_{trans} = \frac{1}{k} \sum_{i=1}^k \frac{\|v_i - \langle v \rangle\|}{\|\langle v \rangle\|} \quad (3.22)$$

where $\| \cdot \|$ is the euclidean length. From the above equation if the time series is deterministic with no added noise $e_{\text{trans}} = 0$ and $e_{\text{trans}} = 1$ if the time series is gaussian noise with a white spectrum. This is a measure of the spread in the displacements of the NNs or plainly the variance of the displacements.

Application to Time Series

Results of the application of the state space method are displayed in fig. 3–19. Numerous simulations with a variety of parameters were undertaken. The specific parameters used in fig. 3–19 are : the number of nearest neighbors for all four time series was 20; the number of reference points was 400, picked at random. The time

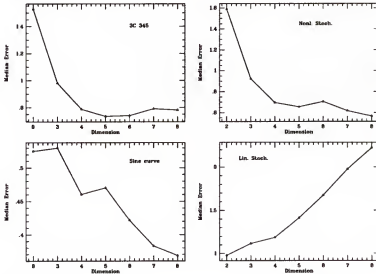


Figure 3–19: The state space continuity method applied to the four time series.

delay for the sine curve, linear stochastic, 3C 345 and NSP time series were 15, 10, 64 and 15 respectively. The time delay for the sin-cos time series was chosen from the first minimum of the autocorrelation function, whereas for the nonlinear time series, 3C 345

and NSP, the first minimum of the redundancy was chosen; this choice of the redundancy over the autocorrelation function is more appropriate for nonlinear time series (Fraser and Swinney, 1986). Nearest neighbors smaller the first minimum of the autocorrelation or redundancy were not chosen in order to avoid artificial correlations (Theiler, 1986).

As is plainly visible from fig. 3–19 the median error for the sine curve time series is much smaller than that for the linear stochastic process. As far as the 3C 345 time series is concerned there is a general consensus that it is not a linear stochastic process (Provenzale et al., 1994, Nair and Smith, 1995). Marked differences can be seen in the redundancy curves (figs. 3–5 bottom left and 3–5 bottom right). Therefore, the issue to consider now is whether the 3C 345 time series is consistent with originating from a process similar to the NSP. From fig. 3–19 it seems that there is almost no difference between the median errors, implying that it may originate from a NSP.

Local Linear Approximation Method

Given a scalar time series $\{x_i\} = x_1, x_2, \dots, x_N$, a reconstructed attractor may be obtained by the method of time delay (Takens, 1981)

$$X(t) = (x(t), x(t - \tau), \dots, x(t - (m - 1)\tau)) \quad (3.23)$$

where τ is the time delay and m the embedding dimension. The points on this reconstructed attractor may be treated as conforming to a map f whose exact functional form is unknown. Now, we assume that the map f is linear in a small region around an attractor point x_i , therefore

$$X_{i+1} = f(X_i) \approx A_i X_i + B_i \quad (3.24)$$

the unknown function is approximated by picking a variable number of neighbors around a chosen reference point. Fit a first order polynomial to these nearest neighbors (NNs); that is, do a linear regression. Higher order polynomials do not seem to produce results that are significantly better than the first order (Farmer and Sidorowich, 1987). NNs less than a certain correlation time are discarded in order to avoid artificial correlations. The number of NNs may be varied between $2(m+1)$ and $N - (m-1) - T$.

We used the popular singular-value decomposition (Press et al. 1992) least squares method to obtain the parameters for the linear fit. The prediction proceeds thus : given the parameters from the least squares fit use eqn. 8 to obtain a value for the future iterate of x_i . Compute the prediction error as the difference between the true value and the computed value.

Application to time series

The local linear approximation does differentiate between the sine curve time series and the linear stochastic process as indicated by the lower error for dimensions 2 through 8 (fig. 3–20). Note that for all dimensions, 2 through 8, there is a clear distinction, in terms of the prediction error, between the two processes. Note also that it is tolerant to

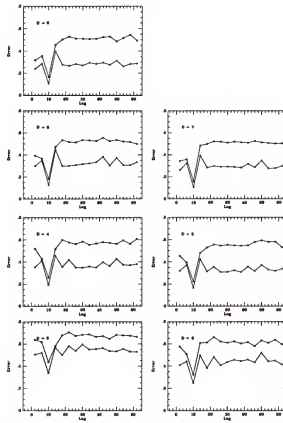


Figure 3-20: Comparison of the prediction error for sine curve and the linear stochastic process using the Local Linear Approximation. In all the boxes the lower is for the sine curve.

high levels of noise. There is a clear difference in the behavior of the errors as a function of dimension for the 3C 345 time series and the NSP (fig. 3-21). As the dimension

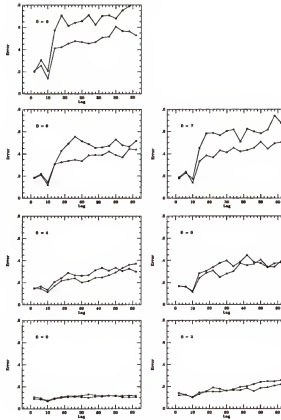


Figure 3-21: The prediction error for the NSP and 3C 345. The curves for 3C 345 are the lower ones.

increases the errors for the 3C 345 time series do not increase as rapidly as that for the NSP. For low dimensions the errors for the NSP and 3C 345 are approximately the same. For higher dimensions the algorithm does resolve the difference between the two processes. At low dimensions the neighboring points in state space will appear to be projected near each other as an artifact due to the small dimension. At large dimensions the attractor will be projected into its true spatial dimension and will thus be resolved (Kennel et al., 1992). This allows the local linear approximation to fit the polynomial more accurately at higher dimensions than at lower ones. In fig. 3-21 for dimensions 6

and greater there is a clear distinction in the errors between the NSP and 3C 345. This implies that there is more predictability in the 3C 345 time series than in the NSP. A similar result was obtained by Nair and Smith (1995) using a zeroth order approximation; that is, using a zeroth order (a constant) fit instead of the first order (linear) fit used here. This is a strong indication that the 3C 345 time series may not be consistent with this NSP.

DISCUSSION AND CONCLUSION

The methods discussed above can be used as part of an arsenal with which to probe all time series, especially experimental ones. The redundancy based tests have clearly shown that both the 3C 345 and the NSP time series are nonlinear and indistinguishable from each other. The state space continuity method too does not differentiate between the two time series either. The “test of noise” algorithm can distinguish between a linear stochastic process and deterministic process on the basis of short-term nonlinear predictability. Nonstationarity of the time series and data truncation do not have any significant effect on the performance of the algorithm. If the input time series is a NSP the algorithm may fail and one needs to interpret the results carefully before arriving at any conclusions. Although the issue of whether the 3C 345 time series is deterministic is obscured by the fact that the algorithm recognizes the NSP from equation 11 as deterministic there are, however, substantial differences between the behavior of the z statistic as a function of the embedding dimension for the two time series; for 3C 345 the mean z statistic indicates a minimum value for a dimension of 4, whereas for the NSP there is no such clear minimum. The “test of noise” algorithm suggests that the 3C 345 time series is not consistent with the NSP generated from eqn. 5–11. Also, the mean

z statistic is lower for 3C 345 than for the NSP. Nonlinear prediction is one the most powerful techniques that exists currently and has fared extremely well in predicting time series (Sauer, 1994). In order to answer the question of whether a time series arises from a stochastic or deterministic process with more certainty and accuracy, one can explore the differences in predictive capabilities of time series using a first order scheme; that is, to fit a linear map to the nearest neighbors -- the Local Linear approximation. The Local Linear approximation does clearly distinguish between the two processes at high dimensions and confirms that the two processes are not the same.

From the application of the "test of noise" algorithm we can see that both time series, 3C 345 and the NSP, are not linear stochastic. Now, the Local Linear approximation method tells us that the 3C 345 time series is not a NSP, as has been claimed by Provenzale et al. (1994). Therefore, we can eliminate the possibility that 3C 345 is a stochastic process, either linear or nonlinear; the alternative is that the 3C 345 time series is from a deterministic process. Fitting higher order polynomials to the nearest neighbors does improve the accuracy of the results (Farmer and Sidorowich, 1987). As Casdagli (1991) mentions, deterministic low dimensional systems can be misidentified as stochastic with small to moderate amounts of data. This may be the most likely scenario in the case of 3C 345. Large amounts of data may required to uncover the rules that govern the system. This does not mean that we should not attempt such analyses. Casdagli (1994) has suggested varying the number of nearest neighbors -- the so-called Deterministic versus Stochastic algorithm (DVS). The basis of the DVS algorithm is that small neighborhoods (small number of nearest neighbors) correspond

to deterministic models, intermediate neighborhoods to nonlinear stochastic models and large neighborhoods to stochastic models. Casdagli has successfully applied the DVS algorithm to a large number of time series. There are, however, a couple of problems with this approach. One is that there is no firm theoretical basis for the assumptions and the second is that this test may be prone to producing erroneous results as suggested by Theiler et al. (1994).

Theiler et al. (1992) have argued for conducting as many tests as possible before arriving at a conclusion about the nature of a time series. The motivation for conducting numerous test on a time series is illustrated in the case of the sunspot time series cited by Theiler et al. (1992). Several authors (Tong & Lim, 1980; Weigend et al. 1990 to name a few) agree that there exists dynamical nonlinear structure – – that is, nonlinear structure originating from the physics – – in sunspot data. In the paper by Theiler et al. (1992) surrogate data methods provide no evidence for nonlinearity whereas using a forecasting algorithm points to evidence for dynamical nonlinear structure. The point is that a certain type of statistic may be more sensitive to a certain type of problem and that one has to examine a range of statistics before arriving at conclusions as to the type of phenomenon one is dealing with. Clearly, in this case of the sunspot data the surrogate data method is inadequate in detecting the underlying nonlinear structure. To quote Theiler et al. (1992) "Just because the original and surrogate time series have the same value for the discriminating statistic, that does not imply that they have the same underlying dynamics"; in the case of Provenzale et al. (1994) that statistic would be the correlation dimension.

The above analyses seem to indicate a dimension of about 4 for the radio-loud quasar 3C 345, indicating that the minimum dimension m_0 into which its attractor may be placed is about 5. Thus one would need at least 5 phase-space variables to characterize the attractor or in other words, five differential equations are needed to describe the evolution of the flux from the source. Surprisingly, this value is not very different from the one that Lehto et al. (1993) obtained for the quite different object NGC 4051. The consistency with which the methods homed in around a value of about 4 makes us believe that this value is indicative of the dimension of the attractor of 3C 345. Such calculations may be worthwhile for the numerous AGN for which there is a reasonable amount of data.

From the application of the algorithm that tests for noise, the 3C 345 system appears to have a fairly large noise content. We have also shown that the light curve of 3C 345 almost certainly has a deterministic component in it. It has been suggested (Gopal-Krishna et al., 1993 and Mangalam and Wiita, 1993) that radio-loud quasars are associated with jets, whereas radio-quiet quasars are not. Babadzhanyants and Belokon (1985) were the first to have proposed a correlation between the birth of superluminal knots and long-term optical variability in 3C 345. Recently, this phenomenon has been observed in another OVV quasar, 3C 279; the large optical flare in 1992 may be linked to the birth of a superluminal radio knot (Wehrle et al. 1994). Ulrich (1993) believes that the UV flux variations of AGN are rapid enough not to be caused by variations in the rate of accretion on the accretion disk. Since the variable optical flux in AGN is thought to have the same origin as the UV, the optical flux may not arise from an accretion disk. If this variation is then connected with jets, the above dimension is indicative of the number of

jet parameters required to completely describe the dynamics of the variations; that is, the number of differential equations required to describe the evolution of the flux is equal to the number of dimensions.

Even though we believe that the time series originates from a low-dimensional (4–5) deterministic system, we do not believe the system is chaotic. Low dimensionality does not directly indicate the existence of chaos. Preliminary computation of lyapunov exponents for the 3C 345 time series produces negative exponents, the largest of which is in the vicinity of -1.0 . Obviously this tends to rule out the possibility that the system may be chaotic.

CHAPTER 4

DYNAMIC MODELING OF TIME SERIES USING ARTIFICIAL NEURAL NETWORKS

The concepts of Artificial Neural Networks (ANNs) have been in use for over two decades now. However, their advent into astronomy as a data analysis tool has been limited to a few applications; for example Rogers and Riess (1994) have applied it to find and classify defects on CCDs. Serra-Ricart et al. 1993 have propounded its benefits to the 2- μ Galactic Survey and Lahav et al. (1994) have used it for pattern recognition tasks in the form of automatic galaxy classification. ANNs have also been used for predicting the yearly mean sunspot numbers (Weigend et al., 1992 and Calvo et al., 1995).

The attraction of ANNs for temporal processing stems from their ability to learn complex nonlinear patterns. The often-used and popular applications of ANNs are in the areas of pattern classification and time series prediction. ANNs are particularly suited for signal processing since they can recognize patterns corrupted by noise. For applications which are time dependent one needs to find a way to represent time in the ANN. This can effectively be accomplished by providing the ANN with dynamic properties to make it responsive to time-varying signals. A static network can be transformed to a dynamic network through recurrent connections or by the additions of memory. Recurrent neural networks are difficult to train and control, so here we will focus on topologies with memory. The most popular approach introduces time delays into the ANN architecture, the time-delay neural network (TDNN) (Lang and Hinton, 1988). However, a more

powerful approach is the gamma model (de Vries and Principe, 1992) where the delay operator is substituted by a general recursive linear operator. The gamma model is capable of finding the delay that minimizes the output mean square error. Here the gamma model will be utilized and compared to the TDNN. Such an ANN may be appropriate to predict the future values to describe the underlying dynamics that generates the time series which can be used to model the underlying physics of the observations. ANNs have been used to predict chaotic signals since 1987 (Lapedes and Farber, 1987). Recently Principe and Kuo (1994) showed better predictive models can be achieved if iterated prediction is used instead of the conventional single step prediction. We will use this methodology here.

We analyze two time series, the first of which is the Whole Earth Observations of the White Dwarf star PG1159-035 described by Clemens (1994). The second is the B-band light curve of the optically violent quasar 3C 345 that has been described by Nair and Smith (1995). The 3C 345 time series is known to be nonlinear (Vio et al., 1991) and is most probably deterministic (Nair and Smith, 1995). A large number of time series in astronomy are now known to be nonlinear, therefore it makes linear analysis of these systems quite inadequate. ANNs can be used to identify the dynamics of the autonomous nonlinear dynamical system that produces the time series. The theoretical basis is the fact that predicting the next point in the reconstructed state space identifies the mapping that defines the underlying dynamical system (Casdagli, 1989). This means that the predicted model when iterated should produce a time series with the same dynamical invariants (correlation dimension and lyapunov exponents) as the original time series. This model can also be used as a predictor of future behavior when seeded with the most recent

samples of the time series. The appealing characteristic of this ANN approach is that the system learns the dynamics directly from the time series when the topological and learning parameters are properly set. After the ANN has learnt the dynamics of the system, it can be used as an autonomous system to generate a new time series that has all the properties of the one it was trained on. Therefore, the ANN can be used to filter the noise and generate a noise-free time series, to fill in missing data, to predict future values of the time series and to generate parametric differential equations as illustrated by Breeden and Hubler (1990). This study is restricted to testing whether the ANN is capable of capturing the dynamics of the system. Here we will be using a focused gamma network and the TDNN as topologies and will be using back propagation through time (BPTT) with teacher forcing to train the dynamic ANN. A commercially available simulator (NeuroSolutions, 1994, see fig. 4-1) was used throughout this study. Finally, ANNs operate in the time domain only. This is a highly desirable property since all underlying frequencies are left unchanged and the time series is not altered in any way.

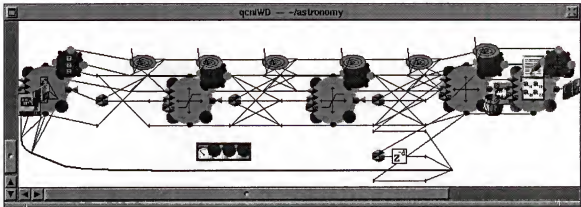


Figure 4-1: The NeuroSolutions bread board with the TDNN setup.

Dynamic Neural Networks

An illustration of the focused dynamic architecture is given in fig. 4-2. It consists of a memory layer followed by a multilayer perceptron (MLP). The TDNN employs a one sample delay operator for G . For the gamma memory

$$G = \frac{\mu}{z^{-1} - \mu} \quad (4.1)$$

where μ is an adaptive constant and z^{-1} is the unit delay operator. The time delay layer implements a time delay embedding (Takens) with a delay of 1 sample. It is therefore crucial to properly assign a dimension for the size of this layer. Takens' theorem provides a lower bound for the size, that is $N > 2m + 1$, where m is the dimension of the attractor. Sophisticated bounds have been proposed by Jones (1990) Kuo and Principe (1994). Notice that the gamma memory allows the use of delays greater than 1. We initially selected the size of the memory . Our final architecture uses a modified gamma topology. Here 3 parallel memory structures with preset μ s have been utilized. The rationale for such a choice is the following. Prediction seeks to find the trajectory on the attractor. Therefore the size of the input memory should span a time length where these features are present. The gamma memory is particularly appropriate to define the time scales due to the free parameter μ . The same number of taps (an input unit) can cover different time spans D by selecting μ from 0.1 to 1.0 according to the formula (Principe et al., 1993) $D = \frac{K}{\mu}$.

Since the gamma memory is also a low pass filter, given by

$$x_k(n) = \mu x_{k-1}(n-1) + (1 - \mu)x_k(n-1) \quad (4.2)$$

changing μ modifies the cutoff frequency of the filter. For small μ (long time scale) the cutoff is moved to zero allowing only the low frequency signal features. We utilized the spectral peaks to help us determine the most appropriate value for μ . In this study the value of N was held fixed at 3 and μ took on values of 0.2, 0.5 and 1.0. We can also interpret this front end as a way to capture features of the time series at multiple scales. We can show that a parallel bank of gamma memories provides a wavelet decomposition (Principe et al. 1995).

Notice that each processing element of the first hidden layer “sees” a single point of the state space trajectory. So the function of the MLP is to find the mapping from a state space trajectory to the next point of the time series. This implements exactly the predictive mapping required for dynamic modeling (Principe and Kuo, 1994).

The input layer consists of neurons that accept time delayed versions of the time series as input. The output of this layer, that is the time-delayed time series, is then passed on to the hidden layer. The strength of the connection between the layers is represented by variable coefficients called weights. The weights are adjusted during the training phase only. The output of one layer feeds all neurons of the next layer.

The output layer usually consists of a single neuron that linearly combines the output from the hidden layer and produces a single output. Fig. 4-2 depicts a basic scheme employed throughout the simulations.

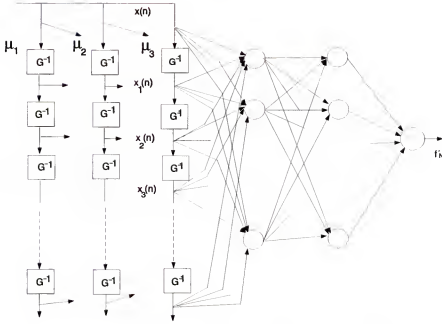


Figure 4-2: A schematic diagram of the Gamma NN

The inputs into a hidden neuron are linearly combined and a bias is added. This can be written as

$$y_i = \sum_{i=1}^m w_{ji} x_i + b_i \quad (4.3)$$

where the $x_1, x_2, \dots, x_j$ are the input signals, $w_{i1}, w_{i2}, \dots, w_{ip}$ are the synaptic weights and b_i is a threshold or bias. The output of each neuron in the hidden layer is bound by an activation function, that is the output of the neuron is

$$x_i(t) = \sigma_i(y_i) \quad (4.4)$$

where σ is the activation function; in this study $\sigma(y) = \tanh(y)$. The tanh activation function performs a continuous mapping in the range $(-\infty, +\infty) \rightarrow (-1, +1)$.

The hidden units or neurons of the hidden layers play a key role in the training process. The hidden units extract the prominent features of the input pattern. Therefore,

it is necessary to use an adequate number of hidden units; using few hidden units will result in the network not capturing all of the features of the input. Whereas, using a large number of hidden units will result in the network modeling the noise present in the input. Our method for choosing the optimum number of hidden layers and units is discussed below.

The primary concern is to try to construct a map from a given time series to study the underlying dynamics. In this study, predicting the future values of the time series is not a major concern. The two tasks, however, are related and as we shall see they are not completely independent of each other. There are two phases involved in system identification from a given time series; the first phase is the training phase and the second is the testing phase. The aim of the network is, given a set of r past samples $x(n-1)$, $x(n-2)$, ..., $x(n-r)$ predict $x(n)$ by building a nonlinear map f , thus

$$x'(n) = f(x(n-1), x(n-2), \dots, x(n-r)) + e(n) \quad (4.5)$$

where $e(n)$ is a white Gaussian process. The computed sample value $x'(n)$ may be compared with the actual value $x(n)$. Therefore, train the net in such a way so as to minimize the prediction error $e(n) = x(n) - x'(n)$. We say that the network has captured the dynamics of the system if the trained network produces a time series that has the same dynamical invariants as the original. In order to test the trained network we run it as an autonomous dynamical system; that is, we freeze the trained weights and feed the output of the trained model as its input (Principe et al. 1992).

Training

The purpose in training is to compute a set of optimal weights to minimize the performance criterion. We use supervised learning to train the network where a teacher signal provides information about how the system should change to improve performance. This information is then embodied in the corrections to be made to the synaptic weights. The performance measure for a supervised learning system is defined on the basis of comparing the desired response or targets and the actual output using a known error criterion. In this study we utilized the BPTT (Werbos, 1990, Williams and Zipser, 1989) algorithm to update the weights with error information over time. The reason for this is twofold. First the focused gamma architecture is recurrent, so a static learning procedure cannot be utilized for good results. Second, we are going to train the ANN as an iterated predictor.

Iterated prediction was shown to constrain the iterates of the mapping, which is desirable in dynamic modeling (Principe and Kuo, 1994). The error is defined as

$$E = \sum_{i=2m+1}^k (x(i+1) - \tilde{x}(i+1))^2 \quad (4.6)$$

where k is the number of iterated predictions and $\tilde{x}(i+1)$ is the predicted output given by

$$\tilde{x}(i+1) = f(x(i), \dots, x(i-2m)) \quad (4.7)$$

This is different from the more conventional MSE used in single step prediction, where $\tilde{x}(i+1)$ is substituted by the output of the system with the previous output. Iterated prediction has been shown to be better than single step prediction (Principe and Kuo, 1994) because the latter may yield a biased estimate. Moreover, the longer the number

of prediction steps are the more are constraints improved on the iterated model which mimics the way the time series is generated.

There are a few issues associated with this training procedure. The first is the choice of topology. In order to implement prediction the output of the net must be fed back to the input. So effectively we are using for training a recurrent system. This system is seeded with N samples from the time series and the output fed back to the input for k samples. It is obvious that if the initial weights are off, the predicted output will be very different from the desired response and the system will take a long time to train (learn). This problem has been recognized before (Williams and Zipser, 1989) and teacher forcing was proposed as a way to speed up training.

Teacher forcing basically helps the ANN to learn faster by inputting segments of the trajectory (instead of a single point), to allow the weights to take reasonable values and then let the system be free for k samples. This procedure is repeated throughout training. The segment length of the forcing is an experimental parameter as well as the number of iterated steps. A basic scheme of this method is illustrated in fig. 4–3.

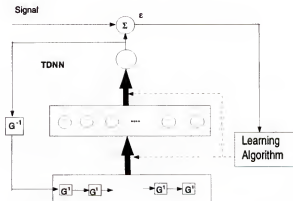


Figure 4–3: A scheme implementing teacher forcing.

During the training the weights are adapted using BPTT. The parameters of BPTT are associated to the size of the trajectory (k) and of course the learning rates and for our case also the momentum rates. A large value for the learning rate helps escape local minima. However, a learning rate that is too large will produce oscillations or instabilities in the ANN. For the tapped delay line memory the learning rate used was 0.4 with a step size of 0.005, whereas for the gamma memory the learning rate was 0.8 with a step size of 0.001. Initially all of the interconnection weights are randomized. Quite naturally the error starts out at high values and as the ANN begins to learn the error drops. We have found, on the basis of the characteristics of the autonomous time series, that a good time to stop training is not when the error has fallen to very small values but to some intermediate values. At very small values of the error the ANN begins to model the noise.

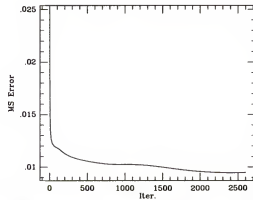


Figure 4-4: The mean square error as a function of the number of iterations for the White Dwarf time series using the tapped delay-line memory. Training was stopped at 2600 iterations.

Fig. 4-4 shows the behavior of the mean-square error as a function of iterations for the White Dwarf time series. Not unexpectedly the error drops steeply within the first 100 iterations and then tries to find a global minimum. The error does show a further

drop with iterations. The criterion for stopping the iterations is based on the following consideration. Initially as the iterations begin the error takes on a large value and drops steadily as the ANN begins to learn. A quantitative criterion for terminating the iterations does not exist. As a result iterations were halted frequently to examine the characteristics of the autonomous time series. It was found that continuing the iterations further leads to overtraining; that is, the ANN models the noise.

Results

White Dwarf time series

The data set for the pre-white dwarf star PG1159-035 was obtained from the Santa Fe Institute (Weigend and Gershenfeld, 1994). Of the 2568 data points in this segment, we used the first 1000 points for training, the 1000 points are depicted in fig. 4-6. As can be seen in fig. 4-6 the time series is contaminated with a significant amount of noise. The power spectrum (fig. 4-5) indicates a dominant period of about 40 time units and two others at around 100 and 250 time units. The high frequency regime beyond $\log(w)$ is not shown in fig. 4-6. The time series originates from a low dimensional system as

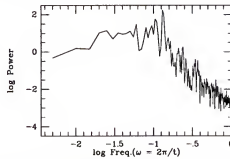


Figure 4-5: Power spectrum of the input white dwarf data.

can be seen from the plot of the correlation integral (fig. 4–8). The correlation dimension computed from the scaling region of the plot, that is for $-2 < \log(r) < -0.8$, is 3.06. It is believed that the variability arises as a result of the passage of the core energy through three separate regions of varying temperature (Clemens, 1994).

The aim is to select reasonable values for the network parameters so as to obtain an ideal fit during autonomous prediction. The parameters listed below were obtained heuristically. A discussion of the various parameters may be in order. Three parallel memory structures were selected to match 3 different time scales as discussed above. The second important parameter is the memory depth, $D = K/\mu$. The choice for K , the number of taps, is related to both the embedding dimension and the delay that provides the best reconstruction. For this data set the embedding dimension is 3. The minimum of the auto correlation is 15. So this means that, in principle, the input signal window should be at least $(2*3+1)*L$. From this input window we only need to take 7 inputs equally spaced by 15 samples to the processing elements (PE). This reduces the number of weights and we call it a sparse connection. The final configuration was a compromise of this reasoning with the goal of providing the time scales to represent each of the major modes of the data. Our system has K equal to 20 taps, each spaced by 2 delays, and a sparse connection to the first hidden layer such that each hidden PE is connected to 5 taps equally spaced in the input signal. With K equal to 40 (20×2), the values of μ to capture the 3 modes are 0.2, 0.5 and 1.

We experimented with two topologies : a simple TDNN with one μ and the Gamma model as described above. One of the problems with the gradient descent technique for

learning is that it may get trapped in local minima. In order to avoid this problem and help the algorithm escape local minima we have a large learning rate. The learning rate and its step size are crucial parameters to assist the network learn the input pattern.

Short-term and long-term prediction errors have been shown by Principe et al. (1992) to be an inappropriate criterion to determine whether or the ANN has learnt a given system. Therefore, Principe et al. (1992) have proposed generating a time series by running the ANN as an autonomous system and examining the dynamical invariants (correlation dimension and lyapunov exponents) of this time series. An autonomous system is one in which the weights are kept unchanged throughout the simulation with the output fed back to the input recursively. The criterion adopted in this study is the same, meaning, we say that the ANN has learnt the system if and only if the autonomously generated time series has the same dynamical invariants as the original time series; we call a time series generated in this manner an autonomous time series.

Several simulations were conducted with a large variety of parameters and the optimal choice of parameters was determined heuristically. An adequate network size is needed during the training for the ANN to learn all the relevant features of the time series. However, an unnecessarily large network size will result in overfitting and will lead to modeling the noise. The network consists of three basic stages, the input layer, the hidden layers and the output layer. The choice of the number of units in the input and hidden layers is crucial to learning. For the white dwarf time series simulations were begun with a small number of units in the hidden layers. These parameters had to be modified by trial and error to obtain satisfactory results. The parameters that gave optimum performance

in terms of the characteristics of the autonomous time series can be found in Table 4-1. It was found that increasing the units in the hidden layers led to the network modeling noise. Whereas, too few units in the hidden layers leads to the network not capturing the essential features.

Table 4-1: Parameters used for the results shown in figs. 4-6 and 4-7

TABLE 4-1							
Tapped delay-line memory							
Input	Delay	1 st	2 nd	Momentum	Step	Teacher	Iterations
taps	(samples)	Hidden layer	Hidden layer		size	Forcing (samples)	
50	1	13	5	0.8	0.001	5	2687
Gamma memory(3 parallel structures)							
Input	Delay	1 st	2 nd	Mometum	Step	Teacher	Iterations
taps	(samples)	Hidden layer	Hidden layer		size	forcing (samples)	
20	2	15	5	0.8	0.001	8	5444

Tapped Delay-Line Memory

Fig. 4-6 shows the result of autonomous prediction (AP) using the tapped delay-line memory, with the top-most being the original time series. The first AP time series was generated by using a few initial points from the original time series as the starting guess.

The simulation was run for 3 epochs or 3000 time units with the starting guess for the 2nd and 3rd epochs being the ending values of the previous epoch. It may be observed that most of the features of the original time series have been captured. Even though the simulated time series, generated by AP, does not exactly follow the original time series, it traces out a slightly different trajectory on the attractor.

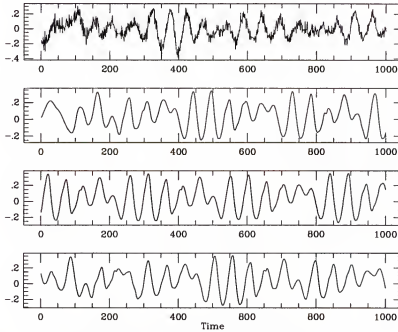


Figure 4-6: Top, the White Dwarf time series. Followed by results of autonomous prediction after 1, 2 and 3 epochs using the tapped delay-line memory.

Gamma Memory

The AP using gamma memory (fig. 4-7) has somewhat more noise than that using the tapped delay-line memory. The top most figure is the original time series followed by the autonomously predicted time series at 3 successive epochs. Apparently some of

the features of the time series seem to be out of phase with the original time series. It does, however, reproduce all of the relevant features of the time series.

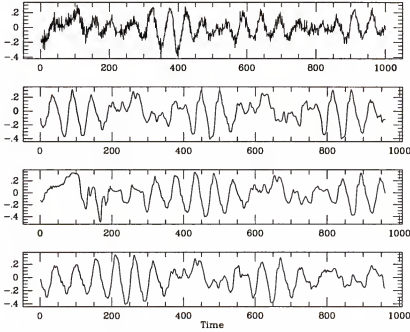


Figure 4-7: The White Dwarf time series, followed by the results of autonomous prediction after 1, 2 and 3 epochs using gamma memory.

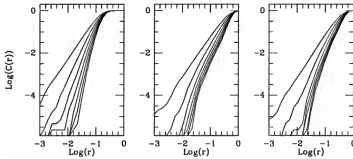


Figure 4-8: Correlation Integral of the White Dwarf time series (left), of the tapped delay-line memory (center) and of the AP time series using the gamma memory (right).

The dimension is a dynamical invariant of any given system. One of the ways of quantifying the dimension is to compute the correlation dimension of the system by

computing the correlation integrals. The correlation dimension is, then, the slope of the curves in the scaling region or in the linear part of the curves. The correlation integrals are shown in fig. 4–8 for dimensions 2 through 8. In the scaling region of fig. 4–8 that is for $-2 < \log(r) < -0.8$ the slopes for the original white dwarf time series are 3.06, the slope computed from the time series generated by AP with gamma memory is 3.00 and the slope for the time series generated using the tapped delay-line memory is 3.23.

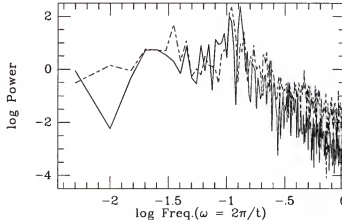


Figure 4–9: Power spectra of the autonomously generated time series. Solid line is the time series from the TDNN and the dashed line from the Gamma NN.

There is a good resemblance between the power spectra of the original time series (fig. 4–5) and that of the AP with gamma memory (dashed line, fig. 4–9). However, the AP with the tapped delay-line memory does seem to be able to reproduce the low frequency structure of the original time series. It is quite evident from these simulations that the ANN is capable of capturing the dynamics of the system. The same methodology is applied to the quasar 3C 345 time series next.

3C 345 time series

The 3C 345 time series consists of sporadic outbursts followed by periods of little or no activity. The occurrence of the outbursts or flares is not periodic. However, one dominant period appears to be present, that is one in the vicinity of about 11.5–12 years (Kidger and de Diego, 1992; Schramm et al., 1993).

After investigating a large number of parameters the tapped delay-line memory did not meet with any success modeling the 3C 345 time series. The autonomous time series either produced low frequency oscillations or other erratic high frequency oscillations. At intermediate values of the learning rate (0.3, 0.4) the autonomous time series was able to reproduce the beginning few segments.

Fig. 4–10 shows an autonomous time series generated using the gamma memory. Clearly the ANN has been able to capture only the low frequency features of the time series and all of the high frequency features have escaped the ANN. This can be explained in terms of the resolution and depth of the gamma memory. The results are quite promising but falls short of our expectancy – it can be seen that the ANN has learnt the some of the low frequency characteristics of the time series. However, the resolution of features is scant and the ANN does not capture the high frequency characteristics.

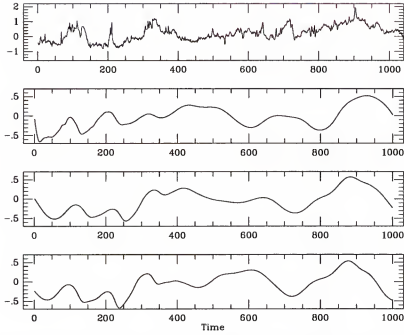


Figure 4-10: Top, the original 3C 345 data. Successively lower figures are simulations using the gamma memory for 3 continuous epochs. The ordinate are normalized flux units.

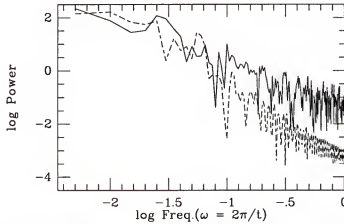


Figure 4-11: Power spectra for 3C 345 (solid line) and the simulated time series.

Fig. 4-11 shows that a fairly good match exists between the power spectra for 3C 345 and the simulated time series at low frequencies but at high frequencies they are markedly

different. Various reasons can be attributed to this partial success, some of which are a) a more sophisticated network may be needed, b) the data may be insufficient, notice that the time series lacks a complete cycle; that is, we have not seen the entire range of activity and have seen only a fraction of all the possible range of activity of the AGN; and c) the system producing the time series may be a nonlinear stochastic system.

Discussion

It is evident from the above results that an ANN has the ability to learn complex data relationships and is a powerful tool for system identification. The gamma memory, which has never before been used to model time series, is clearly more powerful than the extensively-used tapped delay-line memory. Ideally, the ANN requires noise free data with at least one complete cycle or period. However, even in the presence of noise the ANN was able to learn the dynamics of the system as in the White Dwarf time series. In the 3C 345 time series the primary cause for partial success we feel is the lack of a complete cycle. This limitation must be considered before arriving at any conclusions about the system. The trained ANN can be used to fill in missing data and generate a time series with no temporal gaps; it also can be used as a nonlinear filter and produce a time series with almost no noise as can be seen from figs. 4-6 and 4-7. Predicting future values (Calvo et al., 1995) and producing a new time series are further applications of the trained ANN. Parametric differential equations describing the dynamics of the system may be obtained from the new noise-free time series. Finally, successful modeling of the time series is an indication the dynamic system is deterministic, however the reverse

may not be true. The reason being that it may be possible to model the system with a larger and more complex network or a different set of parameters.

CHAPTER 5

AGN TIME SERIES AND STOCHASTIC RESONANCE; A GENERIC MODEL FOR VARIABILITY

Most AGN are thought to be nonlinear and non-stationary systems; the time series of the OVV quasars has been shown to be nonlinear (Vio et al., 1991, Nair, 1995). AGN light curves display a wide range of characteristics and morphologies, meaning that some AGN show frequent, but not strictly periodic, flares (e.g. 1156+295) as shown in fig. 5-1; in others the flares usually do not last for the same amount of time (e.g. 3C 345). Some have shown a large initial outburst followed by a very long period of quiescence (e.g. 2345-167, fig. 5-1). A common feature to all AGN light curves is that the flares, which seem to be superimposed on a noisy background, are of varying amplitudes and morphologies and are unpredictable. Some AGN display multiple peaks during a flare (e.g. 3C 345, 3C 279). Here a generic model is proposed; these features may be explained as a result of SR where noise-induced switching is the crucial ingredient. The most important property of SR is that an optimum amount of noise applied at the input results in coherence at the output. This requirement is met only by a nonlinear system that is also a non-stationary system (Moss, 1992).

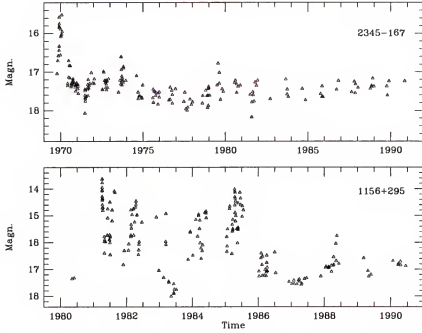


Figure 5-1: Light curves of the OVV quasars 2345-167 (top) and 1156+295 (bottom)

A Physical Model for Variability

The assumption here is that all or at least a major portion of the variability arises due to a jet issuing from the putative central black hole. Births of new knots in radio jets have been associated with flares in the optical and throughout the spectrum on a few occasions, with a finite lag.

The model adopted here is the simple model of Villermaux and Hopfinger (1994). They experimentally simulated a jet issuing forth from a nozzle at velocity u . It was found that for intermediate Reynolds numbers, ranging from 10^3 to about 4000, the jet displays oscillations. The self-sustained oscillations are due to the confinement of the jet. This confined jet is subject to shear (Kelvin-Helmholtz) instabilities, the jet diameter increases as the flow expands in the cavity (or the ISM) and the decrease of velocity

sets up pressure gradients. These pressure gradients give rise to fluid particles at the confining walls and these particles are convected back and join the upstream flow. The interaction of the fluid particles with the jet's shear instability is delayed by time τ . If y is the amplitude of the jet's shear instability, Villermaux and Hopfinger (1994) arrive at an expression for a time evolution of the amplitude as

$$\frac{d}{dt}y(t) = ry(t) - \mu \left(\int_0^{\infty} |y(t-t')|^2 f(t') dt' \right) y(t) \quad (5.1)$$

where r and μ are fixed parameters and $f(t')$ is a general function which modulates the retarded action of the term $y(t-t')$.

Consider a particular simple form of $f(t')$ – the Dirac delta function located at $t' = \tau$; then eqn. 5-1 can be simplified to

$$\frac{d}{dt}y(t) = ry(t) - \mu |y(t-\tau)|^2 y(t) \quad (5.2)$$

Stochastic Resonance

The phenomenon of stochastic resonance (SR) arises as a result of the combined effect of periodic and stochastic forcing in a nonlinear system. SR was first discovered by Benzi et al. (1981) and has since found many interesting applications in many fields. For example, SR has been suggested as a cause for the periodicities observed in the occurrence of the earth's ice ages, in laser experiments (McNamara et al., 1988) and many others.

It is suggested here that SR may be the possible cause for AGN variability. It is intended to show that noise induced switching may occur in the presence of jets and produce similar looking light curves in terms of their broad general features. AGNs

display a wide array of behavior, for example the irregular occurrences of flares, a flare followed by long or short periods of inactivity and almost always an abrupt rise in flux, "crouch before leaping". The appearance of multiple peaks during a flare is common; for example in the case of 3C 345 a "triple-peak" flare occurred around 1990–91. The time series generated by SR is able to display almost all of the features listed above.

Fraix-Burnet (1992) has suggested the existence of stochastic processes in which particles diffuse on magnetic turbulence, this could be the source of noise. Finally, the last ingredient for SR is a low amplitude periodic modulation. Magnetic fields are thought to be present around AGN jets. The periodic modulation can be thought to represent a weak amplitude, low frequency magnetic field entwined perpendicular to the axis of the jet flow. Magnetic fields parallel to the jet flow have a stabilizing effect on the system and inhibit the growth of instabilities and therefore cannot be considered. The above nonlinear differential equation now reads

$$\frac{d}{dt}y(t) = ry(t) - \mu|y(t - \tau)|^2y(t) + c \cdot e(t) + A\cos(\omega t + \phi) \quad (5.3)$$

where $e(t)$ is gaussian noise, c is a multiplying factor, ω is the frequency and A is the amplitude of the periodic modulation. This is a Langevin equation and may be integrated using special techniques employed for stochastic differential equations. For this study a simple fixed step 4th order Runge-Kutta scheme (Press et al. 1992) was used to integrate eqn. 5–3.

The properties of SR may be defined as follows. The physical quantity that characterizes SR is the signal-to-noise ratio; this is defined as the ratio of the amplitude of the signal peak to the amplitude of the noise background in the power spectrum. For

computational convenience the power spectrum is considered only. SR can be observed as the power spectrum rises to a maximum value from low values and goes to low values. The behavior of the time series can be thought of as being in two states, the low state and the high state. The low state is defined/characterized as extended periods of small values for the time series and therefore a high state is defined as extended period of large values for the time series. Switching between states occurs in the presence of noise due to an interaction between the unmodulated system noise and the modulating frequency. When the noise is small the switching rate is almost zero; as the noise increases the interaction increases and there is coherence for an optimum value of the noise. When the noise level is increased further the coherence is lost and the time series is dominated by noise.

The maximum value in the power spectrum occurs at an optimum value for the noise intensity. At this optimum noise intensity the system is driven by noise induced switching that was not available at low noise amplitudes. The introduction of noise induces random switching to different states; the weak periodic signal makes switching coherent with the periodic function (Moss, 1992).

Simulations

Equation 5-3 was numerically integrated using the fixed step size Runge-Kutta routine RK4 (Press et al. 1992). The parameters used were $r = \mu = 0.5$, $\tau = 1$, $A = 0.1$, $\omega = 7.0$ and $\phi = 0.0$. It is customary to examine the effect of increasing the noise variance, c in eqn. 5-3, from very small values to large values. As examples three time series are displayed in fig. 5-2 showing the time series derived from integrating eqn. 5-3.

The time series in fig. 5-2a has a low noise level with variance 0.36, fig. 5-2b has an intermediate noise level of variance 1.04. In fig. 5-2c the noise variance is high, 4.0.

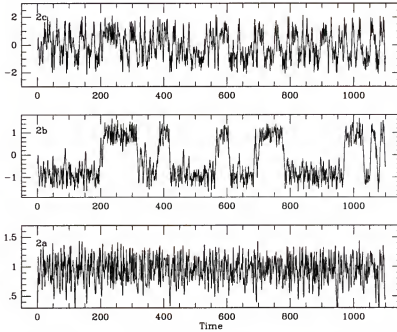


Figure 5-2: Noisy time series generated by integrating eqn. 5-3. The noise variance increases from a) — c). Stochastic resonance can be seen for intermediate noise variance; that is, fig 5-2b.

Clearly, the induced coherence can be seen in fig. 5-2b for an intermediate noise variance. This is depicted in another way in fig. 5-3. The maximum of the power spectra as a function of the noise variance shows an optimum value for the noise variance.

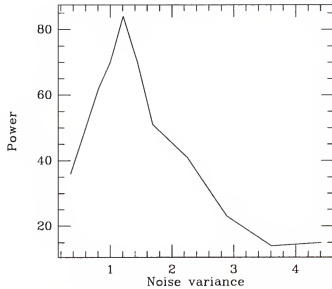


Figure 5-3: Noise variance versus power for the parameters listed above. Indicates an optimum value of 1.2 at which SR occurs.

A variety of light curves can be generated by mainly varying the noise variance and some of the parameters in eqn. 5-3. The system described by eqn. 5-3 is very sensitive to the initial conditions and values of the parameters and as a result it is rare to find two simulated light curves that are alike. As examples consider fig. 5-4, where the value of $\tau = 2$ and all parameters are unchanged from those used to generate the light curves in fig. 5-2. The light curve shown in fig. 5-4a has much lower noise than the one in fig. 5-4b. Fig. 5-4a has some similarity to fig. 5-1 (top) in the sense that the light curve for the quasar 2345-167 shows an initial peak followed by a long period of quiescence. A similar phenomenon can be seen in fig. 5-4a where there is a peak (time 0 to about 200) followed by a period of quiescence. The light curve in fig. 5-4b displays sporadic bursts not unlike that of the quasar light curve, 1156+295, of fig. 5-1 (bottom). However, due to the rapid occurrence of flares in 1156+295 the light curve may be undersampled and

the precise details of the morphology of the flares may be lost. Therefore, effectively the light curve that is observed is the true light curve convolved with the sampling function.

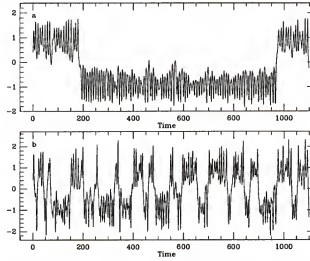


Figure 5-4: Simulated light curves generated by varying the noise variance.

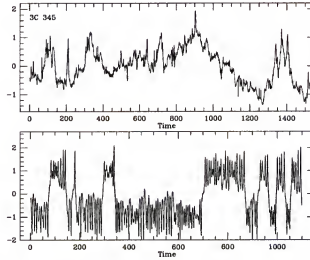


Figure 5-5: Top, light curve of the quasar 3C 345. Bottom, simulated light curve obtained as a result of stochastic resonance.

In fig. 5-5 (top) the 3C 345 light curve extending over roughly 30 years is displayed; in the bottom part of fig. 5-5 a light curve was obtained by integrating eqn. 5-3 with a low level of noise. There is good deal of similarity in the first half of the two light curves. The main feature the simulated light curve lacks and is present in 3C 345 is a secular trend of brightening and fading. This secular trend is not present in most AGN and it is not entirely clear if this trend arises from the sampling of the time series or whether it is intrinsic to the AGN.

In addition microvariability can be explained by this model as originating from random fluctuations of the background flux. These fluctuations are clearly evident in all of the simulated time series. This means that no clear periodicities will be present in all microvariability time series data. Hitherto, no clear case for periodicity have been reported in the literature.

Discussion

Several of the characteristics displayed by AGN light curves can be explained by SR. Some of these are, a general noisy background upon which the flares are superimposed, the noise content in the time series, the irregularity of the occurrence of flares, the occurrence of intermediate sized spikes at irregular intervals and the rapid rise and decline displayed by most AGN flares. SR has the ability to produce arbitrary shapes with changes of noise variance and initial conditions. The difference in individual AGN light curves can be thought of as arising due to differing amounts of noise. Some AGN, e.g. OJ 287, 3C 345, 3C 279 etc., display multiple peaks during flares. The above SR-simulated time series do depict such multiple peaks during an outburst.

Additive noise is the primary distinguishing property for the various AGN time series. According to SR, an AGN time series showing a larger number of flares (e.g. 1156+295) has more additive noise than one with 1 or 2 flares (e.g. 2345-167).

In summary, several predictions can be drawn as a consequence of this model. SR predicts that the system will be nonlinear and nonstationary; this is a property of almost all AGN time series. No two AGN light curves will be alike since all variations are noise driven. The morphology and duration of individual flares will not be identical from flare to flare -- that is, flares of short duration and long duration should be present; also that the occurrence of large flares will not be strictly periodic. Microvariability arises from random background fluctuations and is not periodic. Finally, most AGN time series can be explained as arising from differing levels of noise.

CHAPTER 6 SUMMARY

Summary

I have actively assisted in the AGN monitoring program that has been in progress for over two and a half decades. After switching over from photographic monitoring to CCD monitoring at RHO I have streamlined the operations – specifically the data reduction procedures. The effects of various reductions, e.g. dark currents, flat fields etc., are discussed and their effect on the final magnitude. Many problems can be alleviated with a high signal-to-noise ratio.

Together with the observers at RHO I have participated in almost all major multi-wavelength programs with observers throughout the world. Currently I am participating in a multiwavelength program, that spans all major wavelength regimes from the γ -rays to the radio, to study the variability characteristics of the quasar 3C 454.3 across the spectrum. I have also conducted, with the assistance of the RHO observers, several microvariability runs to study the hitherto unexplored variability characteristics of AGN in the span of minutes, hours and days. These studies then can constrain physical models and will serve as testing grounds for models of AGN variability.

Since many AGN time series are nonlinear, linear techniques are inadequate to analyze these time series. Several nonlinear techniques to analyze time series have been discussed. These include a test for nonlinearity and tests based on nearest neighbors to investigate the nature of a given time series. Numerous authors in the field have

suggested that AGN variability is probably stochastic and arising from a shot noise-like process. In the case of the quasar 3C 345, Provenzale et al. (1994) have suggested that the time series is a nonlinear stochastic process. It has been shown here, for the first time ever, that the 3C 345 time series is nonlinear, deterministic and comes from a relatively low (5–6 dimensions) order system.

Artificial Neural Networks have shown great promise as a data analysis tool and as a system identification tool. While the ANN has performed remarkably well for the White Dwarf time series, it has had modest success with the 3C 345 time series.

Finally, I have suggested a generic physical model for the variability of AGN. The model explains the sudden rapid increase of flux as due to noise-induced transitions. These noise-induced transitions are capable of producing rapid flares and rapid decline of flux. These flares are not strictly periodic and several experiments to test the model have been suggested.

Future Work

Several exciting events, including flares, prolonged minima in the magnitude and change in characteristics of the light curve have been observed for several AGN. All of these behaviors may be trying to give us some information about the character of the underlying mechanism of the AGN. In order to obtain the maximum amount of information from AGN, they have to be monitored as frequently as possible, even at the risk of oversampling. The need for maximum temporal coverage cannot be overemphasized for numerous reasons, not the least of which are to avoid missing flares

and to obtain the morphology of flares. It is always better to monitor a selected few AGN adequately than to monitor several AGN and undersample all, since it will be possible to get a clear understanding of the adequately monitored AGN rather than not have any understanding of any AGN!

Some AGN (BL Lac, 3C 120) change behavior over time and they have most probably not yet exhibited their entire repertoire.

In order to confirm these results a similar analysis can be done on the radio time series if there are adequate data. These analyses must be extended to several other AGN in order to obtain a more comprehensive understanding of the phenomenon of AGN.

The SR resonance model is subject to short-term and long term testing. An experiment that can easily be performed as a short-term prediction test of the SR model is to conduct a microvariability study on AGN that show microvariability. Typically, one should monitor a radio-loud AGN over a certain period to see if any periodicity is present. A large majority, 23 out of 25, radio-quiet AGN have not displayed any microvariability and in most cases of microvariability found in AGN have been restricted to radio-louds (Jang, 1995). Rigorous temporal monitoring of as many flares as possible for a given AGN must be conducted to ascertain the duration of flares and their morphology. SR resonance predicts that no two will be exactly alike.

Several other ANN schemes can be employed to determine whether the ANN can learn the system that produces these time series.

BIBLIOGRAPHY

- Angione, Ronald J., Moore, Elliott P., Roosen, Robert G., Sievers, John, 1981, *Astron. J.*, **86**, 653
- Babadzhanyants, M. K., Belokon, E. T., Semenova, E. V., 1975, *Trudy Astr. Obs.*, **31**, 100
- Babadzhanyants, M. K., Belokon, E. T., 1985, *Afz*, **23**, 459
- Babadzhanyants, M. K., Belokon, E. T., Denisenko, E. T., Semenova, E. V., 1985, *Soviet Astr.*, **29**, 394
- Badii, Remo and Politi, Antonio, 1985, *Jour. of Stat. Phys.*, **40**, 725
- Barbieri, C. and Abati, Erculiani, L., 1968, *Mem. Soc. Astr. Italiana*, **39**, 421
- Barbieri, C. Romano, G., di Serego, A. S., Zambon, M., 1977, *Astron. and Astrophys.*, **59**, 419
- Benzi, R., Sutera, A. and Vulpiani J., 1981, *J. Phy. A*, **14**, L453
- Benzi, R., Parisi, G., Sutera, A. and Vulpiani J., 1982, *Tellus*, **34**, 10
- Boyle, S. A. and Smith, G. Q., 1970, *A. J.*, **65**, 107
- Calvo, R. A., Ceccatto, H. A. and Piacentini, R. D., 1995, *Ap. J.*, **444**, 916
- Casdagli, Martin, 1991, *J. Roy. Stat. Soc. B*, **54**, 303
- Casdagli, Martin C. and Weigend, Andreas S., 1994, *Nonlinear Modeling and Forecasting*, Santa Fe Institute Studies in the Sciences of Complexity, pg. 347, Proc. Vol. XII, eds. Martin Casdagli and Stephen Eubank, Addison-Wesley, Reading, MA.
- Clemens, J. Christopher, 1994, *Nonlinear Modeling and Forecasting*, Santa Fe Institute Studies in the Sciences of Complexity, pg. 175, Proc. Vol. XV, eds

Andreas Weigend and Neil A. Gershenfeld, Addison-Wesley, Reading, MA.

Chatfield, C. and Collins, A. J., 1980, *Introduction to Multivariate Analysis*, Chapman and Hall, London, New York.

de Vries, B. and Principe, J. C., 1992, *Neur. Net*, **5**, 565

Eckmann, J.-P., and Ruelle, D., 1985, *Rev. of Mod. Phys.*, **57**, 617

Farmer, J. Doyne, Ott, Edward and Yorke, James A., 1983, *Physica D*, **7**, 153

Farmer, J. Doyne, and Sidorowich, John J., 1987, *Phy. Rev. Lett.*, **59**, 845

Fraser, Andrew M. and Swinney, Harry L., 1986, *Phy. Rev. A*, **33**, 1134

Garcia, L., Webb, J. R., Gerstman, B. and Nair, A. D., 1991, *B. A. A. S.*, **179**, no. 4, 1421

Gilliland, Ronald L., 1992, *Astronomical CCD Observing and Reduction Techniques*, ed. Steve B. Howell, pg. 68, Astronomical Society of the Pacific, San Francisco.

Gopal-Krishna, Sagar, Ram and Wiita, Paul J., 1993, *M.N.R.A.S.*, **262**, 963

Grassberger, Peter, and Procaccia, Itamar, 1983, *Phy. Rev. Lett.*, **50**, 346

Haykin, Simon., 1994, *Neural Networks*, Macmillan College Publishing Co., New York.

Henden, Arne A. and Kaitchuck, Ronald H., 1990, *Astronomical Photometry*, Van Nostrand Reinhold, New York.

Horne, James H. and Baliunas, Sallie L., 1986, *Ap. J.*, **302**, 757

Jang, Min-Huan, 1995, Ph.D thesis, Georgia State University, Atlanta.

Jones, R. D. et al., 1990, Function approximation and time series prediction with neural networks, *Tech. Rep.*, **LA-UR-90-21**, Los Alamos National Laboratory,

Los Alamos, NM.

Kennel, M. B. and Isabelle, S., 1992, *Phy. Rev. A*, **46**, 3111

Kennel, M. B., Brown, Reggie and Abarbanel, Henry D. I., 1992, *Phy. Rev. A*, **45**, 3403

Kidger, M. R. and de Diego, J. A., 1990, *Astron. and Astrphy.*, **227**, L25

Kidger, M. R. and de Diego, J. A., 1992, *Variability of Blazars*, ed. E. Valtaoja and M. Valtonen, pg. 382

Kinman, T. D., Lamla, E., Ciurla, T., Harlan, E., Wirtanen, C. A., 1968, *Ap. J*, **152**, 357

Lang, K. J. and Hinton, G. E., 1988, The development of the time-delay neural network architecture for speech recognition, Tech. Report CMU-CS-88-152, Carnegie-Mellon Univ., Pittsburgh, PA.

Lapedes, A. and Farber, R., 1987, Nonlinear signal processing using neural networks : Prediction and system modeling, Technical report LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM.

Lehto, H.J., Czerny, Bozena and McHardy, Ian M., 1993, *M.N.R.A.S*, **261**, 125

Llyod, C., 1984, *M.N.R.A.S*, **209**, 697

Lu, Phillip, K., 1972, *A.J.*, **77**, 829

Mangalam, Arun V. and Wiita, Paul J., 1993, *Ap.J.*, **406**, 420

McNamara, B., Wiesenfeld, K. and Roy, R., 1988, *Phy. Rev. Lett.*, **60**, 2626

Moss, Frank, 1992, *Frontiers in Applied Mathematics*, ed. George H. Weiss, SIAM Philadelphia.

Nair, A. D. and Smith, A. G., 1995, *M.N.R.A.S*, under revision

Palus, Milan, 1994, *Nonlinear Modeling and Forecasting*, Santa Fe Institute Studies in the Sciences of Complexity, pg. 387, Proc. Vol. XV, eds. Andreas Weigend and Neil A. Gershenfeld, Addison-Wesley, Reading, MA.

Palus, Milan, Albrecht, Vladimir and Dvorak, Ivan, 1993, *Phy. Rev. A*, **175**, 203

Press, William H., Teukolsy, Saul A., Vetterling, William T. and Flannery, Brian P., 1992, *Numerical Recipes*, 2nd ed., Cambridge Univ. Press., Cambridge, U.K.

Principe, Jose C., Rathie, Alok and Kuo, Jyh-Ming, 1992, *Intl. Jour. of Bifur. and Chaos*, **2**, 989

Principe, Jose C., de Vries, B. and Oliveira, P., 1992, *IEEE Trans. Sig. Proc.*, **41**, 649

Principe, Jose C., and Kuo, Jyh-Ming, 1994, *Advances in Neural Information Processing*, eds. Jack D. Cowan, Gerald Tesauro and Joshua Alspecter, **7**, 311

Principe, Jose C., Jose C., Celebi, Samel, Wang, Chuan, Lefebvre, W. Curt, Harris, John G., and Euliano, Neil R., 1995, *Neur. Net*, in press

Provenzale, A., Smith, L. A., Vio, R. and Murante, G., 1992, *Physica D*, **58**, 31

Provenzale, A., Vio, R. and Cristiani, A., 1992, *Ap. J.*, **428**, 591

Sauer, Tim, 1994, *Nonlinear Modeling and Forecasting*, Santa Fe Institute Studies in the Sciences of Complexity, pg. 175, Proc. Vol. XV, eds. Andreas Weigend and Neil A. Gershenfeld, Addison-Wesley, Reading, MA.

Schramm, K.-J., Borgeest, U., Camenzind, M., Wagner, S. J., Bade, N., Dreissigacker, O., Heidt, J., Hoff, W., Kayser, R., Kuhl, D., Linde, J. v., Linnert, M. D., Pelt, J., Schramm, T., Sillanpaa, A., Takalo, L. O., Valtaoja, E., and Vigotti, M., 1993, *Astron. and Astroph.*, **278**, 391

Serra-Ricart, Miquel; Calbet, Xavier; Garrido, Lluís and Gaitan, Vicens, 1993, *Astron. J.*, **106**, 1685

- Serre, T., Auvergne, M. and Goupil, M. J., 1993, *Astron. and Astrophys.*, **259**, 411
- Shinbrot, Troy, Grebogi, C., Wisdom, J. and Yorke, James A., 1992, *Amer. Jour. Phy.*, **60**, 491
- Simonetti, J.H., Cordes, J.M. and Heeschen, D.S., 1985, *Ap. J.*, **296**, 46
- Smith, A. G., Nair, A. D., and Clements, Sandra D., 1991, *Variability of active galactic nuclei*, Georgia State University Conference on Variability of Active Galactic Nuclei, pg. 52, Cambridge Univ. Press, Cambridge, U.K.
- Smith, A. G., Nair, A. D., Leacock, R. J. and Clements, Sandra D., 1993, *Astron. J.*, **105**, 437
- Smith, Alex G., Nair, A. D., 1995, *P. A. S. P.*, **107**, 863
- Takens, F., 1981, *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics 898, Springer-Verlag, Berlin, 366
- Terrell, J. and Olsen, K. H., 1970, *Ap. J.*, **161**, 399
- Theiler, J., 1986, *Phy. Rev. A*, **34**, 2427
- Theiler, James, Galdrikian, Bryan, Longtin Andre, Eubank, Stephen and Farmer, J. Doyne, 1992, *Nonlinear Modeling and Forecasting*, Santa Fe Institute Studies in the Sciences of Complexity, pg. 163, Proc. Vol. XII, eds. Martin Casdagli and Stephen Eubank, Addison-Wesley, Reading, MA.
- Theiler, James, Linsay, Paul S. and Rubin, David M., 1994, *Nonlinear Modeling and Forecasting*, Santa Fe Institute Studies in the Sciences of Complexity, pg. 429, Proc. Vol. XV, eds. Andreas Weigend and Neil A. Gershenfeld, Addison-Wesley, Reading, MA.
- Tong, H. and Lim, K. S., 1980, *J. Roy. Stat. Soc. B*, **42**, 245
- Tornikoski, M., Valtaoja, E., Terasranta, H., Smith, A. G., Nair, A. D., Clements, S. D., and Leacock, R. J., 1994, *Astron. and Astrophys.*, **289**, 673

Ulrich, M. H. D, 1993, Preprint, 2nd Workshop on Theory of Accretion Disks, Munich, Germany.

Vio, R., Cristiani, S., Lessi, O. and Salvadori, L., 1991, *Ap. J.*, **380**, 351

Vio, R., Cristiani, S., Lessi, O. and Provenzale, A., 1991, *Ap. J.*, **391**, 518

Wayland, Richard; Bromley, David; Pickett, Douglas and Passamante, Anthony, 1993, *Phy. Rev. Lett*, **70**, 580

Webb, J. R, Shrader, C. R., Balonek, T. J., Crenshaw, D. M., Kazanas, D., Clements, S., Smith, A. G., Nair, A. D., Leacock, R. J., Gombola, P. P., Sadun, A., Miller, H. R., Robson, I., Fujimoto, R., Makino, F., Kii, T., Aller, H., Aller, M., Hughes, P., Valtaoja, E., Terasanta, H., Salonen, E., Tornikoski, M. and Chism, W., 1994, *Ap. J.*, **422**, 570

Webb, J. R, Crenshaw, D. M., Shrader, C. R., Balonek, T. J., Clements, S., Smith, A. G., Nair, A. D., Leacock, R. J., Gombola, P. P., Sadun, A., Miller, H. R., 1991, *B.A.A.S*, **26**, no. 4, 1420

Wehrle, A. E, Zook, A. C., Unwin, S. C., Urry, C. M. and Gilmore, D., 1994, *B.A.A.S*, **26**, 1507

Weigend, A., Huberman, B. and Rumelhart, D., 1990, *Int. J. Neural Sys.*, **1**, 193

Weigend, Andreas S., Huberman, Bernardo A. and Rumelhart, David E., 1992, *Nonlinear Modeling and Forecasting*, Santa Fe Institute Studies in the Sciences of Complexity, pg. 395, Proc. Vol. XII, eds. Martin Casdagli and Stephen Eubank, Addison-Wesley, Reading, MA.

Weigend, Andreas S. and Gershenfeld, Neil, E., 1994, ftp.santafe.edu, *Time Series Prediction : Forecasting the future and understanding the past*, Santa Fe Institute Studies in the Sciences of Complexity, pg. 569, Proc. Vol. XV, Addison-Wesley, Reading, MA.

Werbos, P. J., 1990, *Proc. of the IEEE*, **78**, 1550

Williams R. J. and Zipser, D., 1989, *Neural Computation*, **1**, 270

Wing, Robert F., 1973, *Astron. J.*, **78**, 684

APPENDIX B

1.) UV-Optical Observations of the 1991 Outburst of 3C 345, Webb et al. (1991).

Abstract. The blazar 3C 345 underwent a 2.5 magnitude optical outburst in early 1991. The flare was monitored optically in the B and R bands and also in the UV. We present twelve well-exposed LWP and SWP spectra taken at roughly two week intervals during the decline of the burst along with 27 R measurements and 28 interspersed UBVI continuum measurements. We also analyze all of the archival IUE data for this object. We find that the UV spectrum exhibits strong, broad redshifted Lyman alpha and carbon IV lines. We investigate the variability of the UV-optical continuum and the variations of the broad UV lines and find that there is essentially no delay between the UV/optical continuum throughout the observations and that the spectral variations, if present at all, are small. We find no evidence that the Lyman alpha line has varied.

2.) Time Series Analysis of 10 Exosat Observations of PKS 2155-304, Garcia et al., (1991).

Abstract. We have extracted 10 EXOSAT ME observations of PKS 2155-304 and analyzed them using unequal interval Fourier transform and CLEAN techniques and structure functions. We find that in only two of the observations can the variability be unambiguously attributed to $1/f$ noise in the source. We also investigate the effects of sampling and the presence of a white noise component on the reliability of the time series techniques by generating pure $1/f$ noise, adding varying amounts of white noise and then sampling the simulated data identically to the sampling of the x-ray light curves. We find that the sampling is unimportant and that the white noise component masks the $1/f$ character in the high frequency range, but not in the low frequency part of the PDS.

3.) Some Long-Term Properties of a Large Group of AGN, Smith et al., (1991).

Abstract. The analysis of the long-term optical observational data for the database of Active Galactic Nuclei (AGN) currently being monitored at Rosemary Hill observatory is presented. It is found that most AGN display their full amplitude of optical variability within five to ten years. These variations can be broadly categorized as rapid; that is, occurring on time scales of days to months and secular; that is, variations occurring on time scales of years. For most objects the rest frame time scales averaged somewhat over half of the observed time scales. The most common time scale range was three to six years. It is observed that the amplitude of variations seems to decrease as the redshift increases and that the rest frame time scales appear to be shorter at large redshift. The observation of two quite different time scales in many sources strongly suggests that we may be seeing emission from two physically separated regions. Perhaps the impulsive

bursts originate near the central engine, while the long-term changes reflect events in more distant regions.

4.) The Longer Optical Time Scales of a Large Sample of Quasars, Smith et al., (1993).

Abstract. Of a carefully studied sample of 60 quasars, 75% displayed evidence of long-term changes in the base levels of their optical fluxes. Internally consistent light curves were analyzed by three different methods in an effort to assign time scales to these changes. The familiar rapid, large-amplitude flares were specifically excluded from the study. The majority of the observed time scales range from 3 to 15 yr, but when they are translated into the rest frames of the objects, there is a strong concentration between 2 and 6 yr. Extensive testing showed that the lengths of the data runs, largely 17 to 21 yr, did not exert a major influence on the estimation of the time scales. Several accepted source mechanisms appear to be capable of generating variations with appropriate time scales.

5.) Correlated Radio and Optical Variations in a Sample of AGN, Tornikoski et al., (1994).

Abstract. We have studied flux variations in the optical and at radio wavelengths of a sample of extragalactic sources. Even though similar analyses have been done earlier, this is the first extensive study of connections between long-term optical and high radio frequency (from 22 GHz up to 230 GHz) variability in a large sample (22 sources) of AGN. For the correlation analysis we used the discrete correlation function complemented by visual comparison and concentrated on finding correlations with time delays of about 1000 days and less, since any time delays longer than some hundred days are very hard to verify. In 10 out of the 22 sources we found a clear radio-optical correlation, and in as many as 6 sources the optical and high frequency radio variability was simultaneous. We suggest that the main reason for not finding significant correlations in the rest of the sample sources is the lack of sufficient radio and especially optical data.

6.) The Multifrequency Spectral Distributions Evolution of Blazar 3C 345 During the 1991 Outburst, Webb et al., (1994).

Abstract. The blazar 3C 345 underwent a 2.5 mag optical outburst between 1990 November and 1991 May. We have obtained 10 nearly simultaneous multifrequency spectra during the course of the outburst in order to study the multifrequency spectral variations of 3C 345 as a function of time. Although our observations were not sampled frequently enough to completely resolve the variations in every frequency band, the general rise and decline of the outburst were seen in the UV through radio with differing rise times. Simulations of an electron distribution injected into a tangled magnetic field show a relationship between frequency and characteristic timescale that was also observed in the radio variations of 3C 345. The two X-ray observations made during the monitoring

period showed no evidence of variability. The multifrequency spectrum was modeled with two major components : a relativistic jet and a relativistic thermal accretion disk. Models calculated for each spectrum indicate that the outburst can be explained in terms of these models by varying the high-energy cutoff of the injected electron distribution in the jet model, while also varying the mass accretion rate in the disk model. There is marginal evidence that the inferred accretion rate varies with the jet luminosity.

7.) Time Scales of Long-Term Optical Base-Level Fluctuations in Three Classes of AGN, Smith and Nair, (1995).

Abstract. The long-term optical light curves of many active galactic nuclei (AGN) display smooth, cyclic changes in level on time scales of years. Earlier work by Smith et al. (AJ, 105, 437, 1993) estimated the time scales of this previously undiscussed phenomenon for a sample of 60 classical radio-loud quasars. The present study extends the analysis to three additional groups of AGN : the BL Lacertids, the compact galaxies, and the radio-quiet quasars. The average time scale of the BL Lacertids is approximately equal to that of the classical quasars, lending credence to unified models of these objects. The long-term fluctuations of the compact galaxies and the radio-quiet quasars are slower. Our analysis tends to support the division of the BL Lacertids into two populations.

BIOGRAPHICAL SKETCH

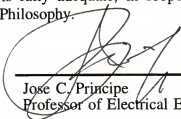
I earned a Bachelor of Engineering in 1983, from Banglore, India and a Master's degree in Physics in 1989 from Bowling Green State Univ., Bowling Green, OH. My primary area of interest has been signal processing, time series analysis and nonlinear dynamics. I intend to and I am apply signal processing techniques to image processing. In Jan. 1993 I added a new dimension and perspective to life by marrying Jayshree.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



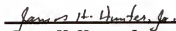
Alex G. Smith, Chair
Distinguished Service
Professor of Astronomy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Jose C. Principe
Professor of Electrical Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



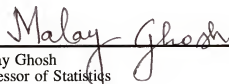
James H. Hunter, Jr.
Professor of Astronomy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Robert J. Leacock
Associate Professor of Astronomy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

A handwritten signature in dark ink, reading "Malay Ghosh", is written over a horizontal line.

Malay Ghosh
Professor of Statistics

This dissertation was submitted to the Graduate Faculty of the Department of Astronomy in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 1995



Dean, Graduate School